Egocentric value maps of the near-body environment: from Reinforcement Learning to neural responses

Rory John Bufacchi[1], Richard Somervail[1], Aoife Maria Fitzpatrick[1], Roberto Caminiti[1], Gian Domenico Iannetti[1,2]

[1]rory.bufacchi@gmail.com, rory.bufacchi@outlook.com, +447590267731

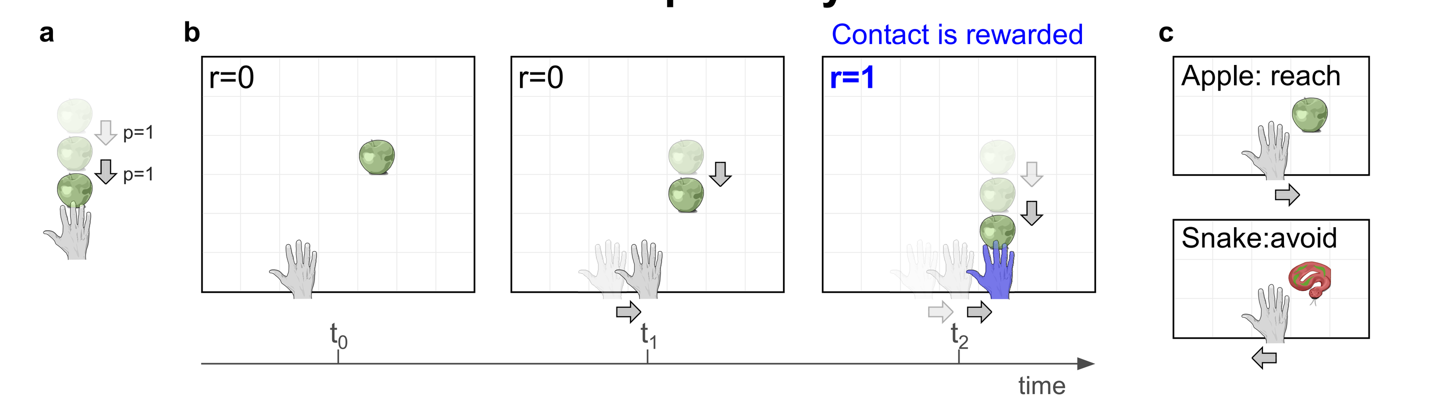
UCL

iit ISTITUTO ITALIANO DI TECNOLOGIA

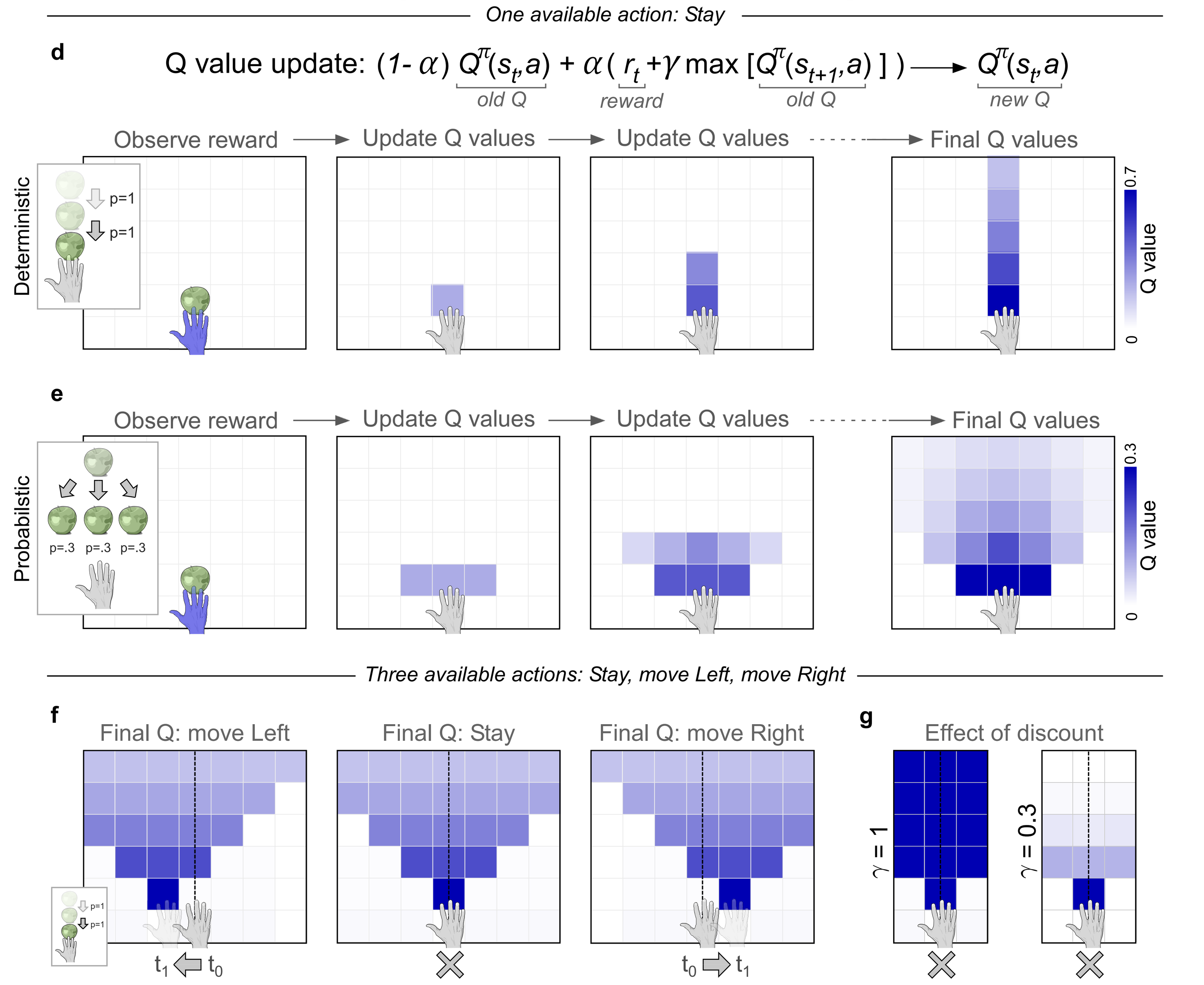
Background: Body-part centric response fields are pervasive, but remain poorly understood because we lack a unifying formal explanation of their origins and role in wider brain function. Here, we provide such explanation.

Methods and Aims: We use reinforcement learning to analytically explain the existence of body-part centric response fields, also known as peripersonal fields. We then simulate multiple experimental findings considered foundational in the peripersonal space literature, and directly fit empirical data from 21 previously published experiments from 8 research groups.

World setup and dynamics



Body-part centred fields emerge



Explanation of theory: Consider a simple environment in which an object always moves on a downward trajectory (**a**), and **contact** between the object and a given body part is **rewarded** (**b**). The agent will move towards objects which increase reward (**c**, top) and away from objects that decrease it (**c**, bottom).

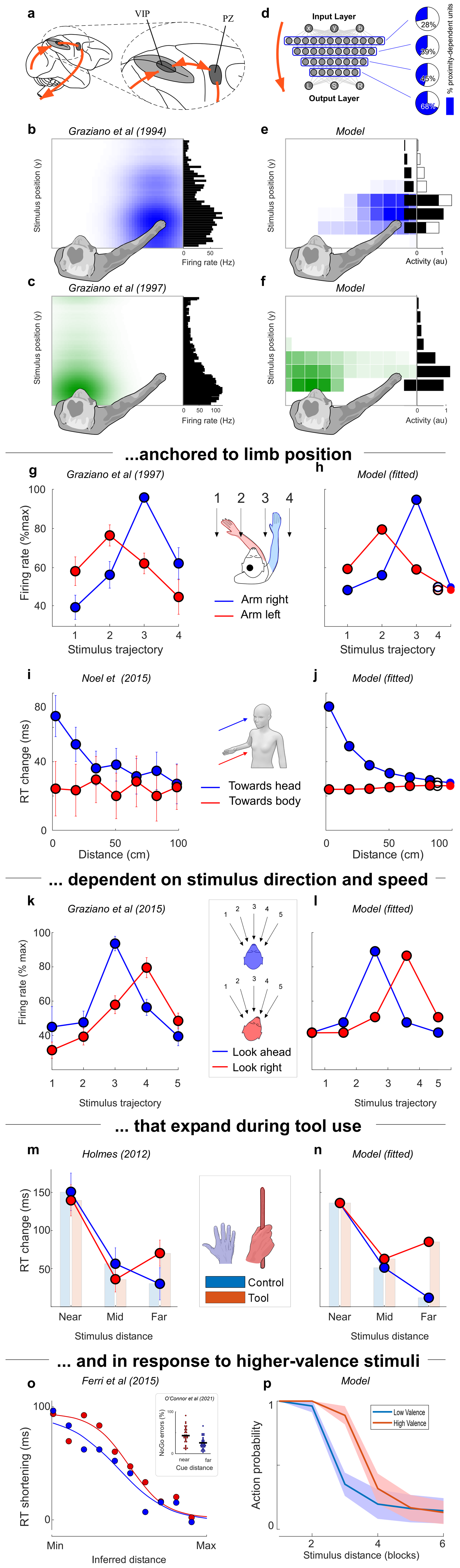
d In Reinforcement learning, the agent chooses its actions by calculating the **value Q of performing those actions**. The agent incrementally updates the value of the 'stay' action, eventually arriving at the optimal value (right). Because staying in place can create or avoid contact, the **value of performing that action will form an egocentric receptive field**, centred around the body part: a **peripersonal field**.

e Under different world dynamics, the value-field will take a different shape.

f The actions available to an agent also affect the positions from which an object can contact the body: **motor repertoire expands peripersonal fields.**

g Temporal discount also creates an inverse relationship between stimulus distance to a body-part and action value.

Underlying neurons have body-part centred response fields



Results: Peripersonal fields

Peripersonal fields naturally emerge from two simple and plausible assumptions :

1) living agents experience reward when they contact objects in the environment

2) they act to maximise reward.

These simple assumptions give rise to egocentric action-value fields that explain empirical findings on stimulus kinematics, tool use, valence, and network-architecture.

Model comparison to data:

a-c Macaque brain areas VIP and PZ house **neurons with body-part centred receptive fields**.

d-f Artificial networks contain similar neurons when trained to simultaneously move two 'body-parts'. Different artificial neurons in respectively have 'limb' and 'face' centred receptive fields. The proportion of neurons with such receptive fields increases as a function of layer depth (**d**).

g The firing rate peak of the neurons with arm-centred receptive fields **moves with limb position**.

h As does the peak of artificial body-part centred neurons. ('fitted' indicates that the model has been numerically fitted to the data)

i Such action-value neurons provide a putative substrate for the many body-part centred behavioural responses observed in humans, as demonstrated by a solid model fit (**j**).

k Canonical biological peripersonal fields depend on **stimulus velocity and direction**.

l Artificial value fields also expand when incoming stimuli move faster and from different directions.

g Canonical peripersonal fields **extend** to incorporate the **tip of a tool**, specifically after training with it (left). Similarly, artificial value fields expand only after training with a tool that increases the ability to touch an object (right)

k Proximity-dependence of peripersonal measures is stronger for stimuli of **higher valence**. Relatedly, stimuli with high valence more frequently elicit spontaneous movements when the stimuli are near (inset).

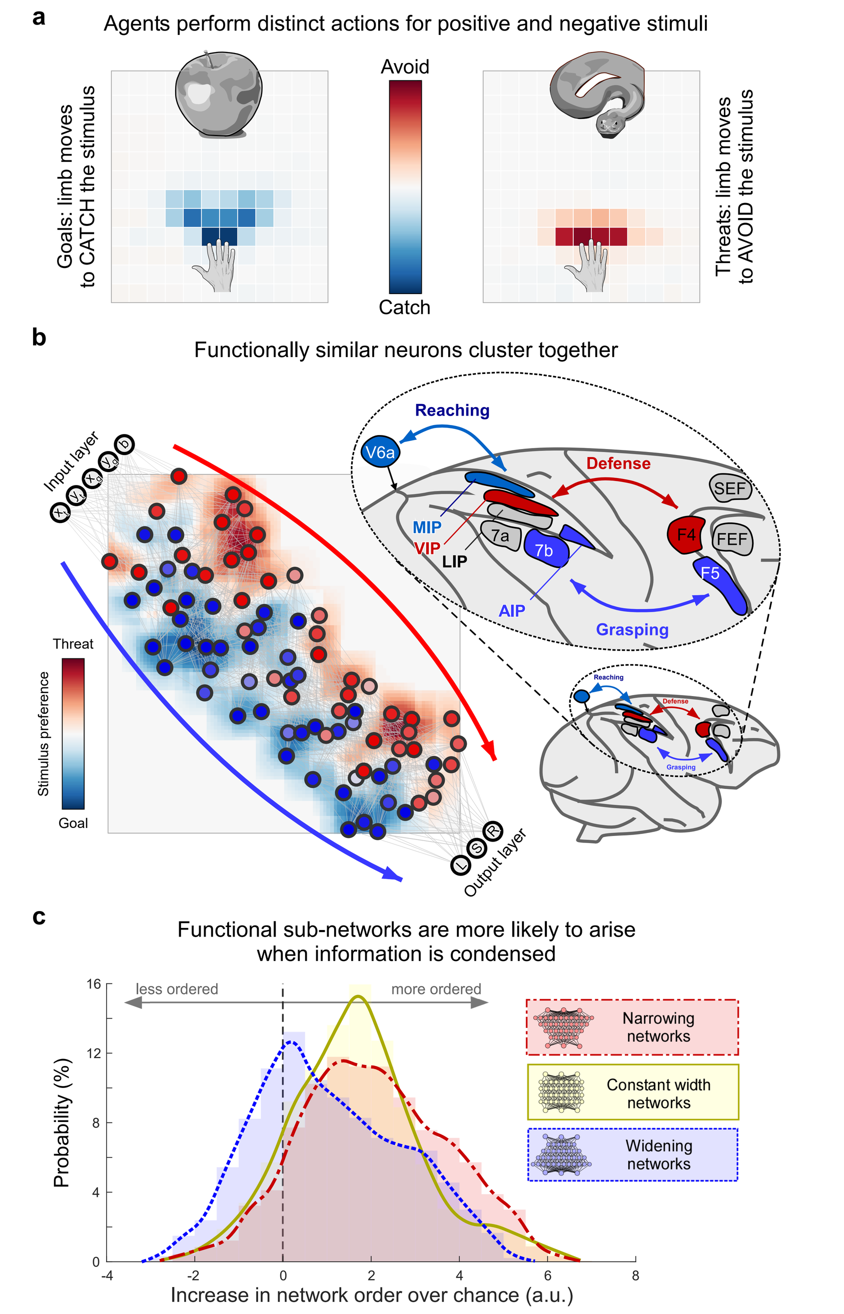
l Accordingly in artificial agents, actions (that aim to create or avoid contact) are initiated at further distances in response to a higher-valence object.

erc
wellcome trust

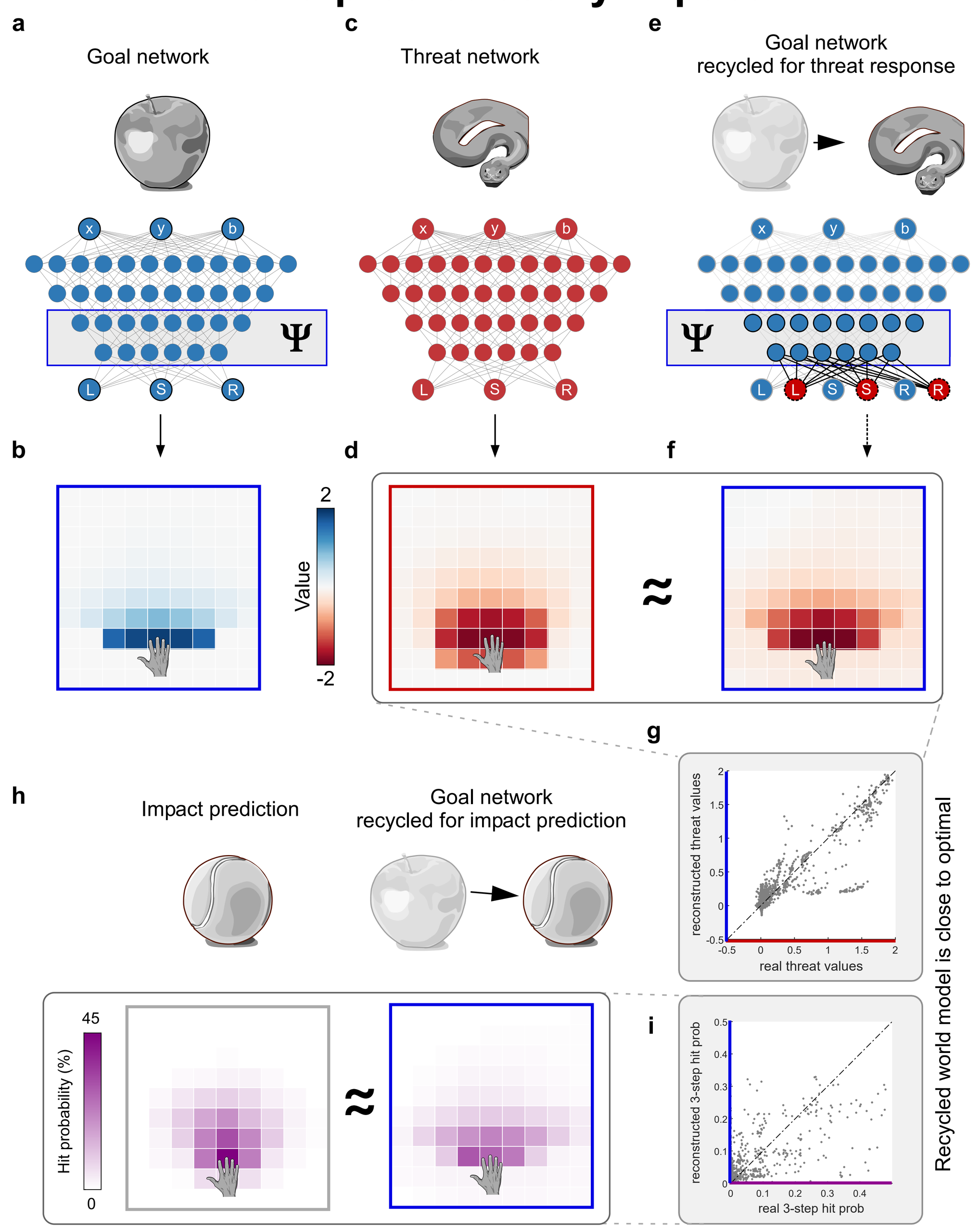
Results: Egocentric maps

Our explanation offers a formal description of the notion that the world-agent state is encoded in parieto-premotor cortices using motor primitives; **peripersonal fields** provide short-term **building blocks** that together create a **map** of the **world near the agent in terms of its future states**: a successor representation. This short-term, close-range egocentric peripersonal map is analogous to the long-term, long-range allocentric spatial map of place and grid cells, which underlie locomotion and navigation to reach distant objects. Together, these allocentric and egocentric maps allow efficient interactions with a changing environment across multiple spatial and temporal scales.

Functionally distinct sub-networks emerge



A set of approximate value fields (Ψ) can be 'recycled' into new value functions and predict bodily impact



Functional sub-networks emerge:

a When trained on positive and negative reward stimuli, artificial agents display different patterns of motor activity.

b Training an artificial network to perform both approach and avoidance behaviors (as in a) gives rise to **spatially distinguishable sub-networks** (red vs blue; network graph on the left). This is reminiscent of the anatomical structure of the parieto-premotor system, where peripersonal neurons cluster together based on their behavioural function (inset on the right).

c, Such sub-network structure is particularly likely to appear when the network condenses information (i.e. when it narrows; pink histogram), compared to when it spreads out information over many neurons in later layers (i.e. when it widens; blue histogram).

Egocentric maps:

Peripersonal fields could be used as **basis functions** to flexibly interact with the world near the body. An artificial network that has only learned to reach positive valence stimuli (**a**,**b**) can be 'recycled' to approximate an appropriate value field for avoidance movements (**c**,**d**). Specifically, by taking a weighted sum of the neural activities in the second half of the blue network PSI, the output from the red network (**e**) could be faithfully reconstructed (**f**,**g**).

h Furthermore, the probability that a stimulus would hit the body over any number of timesteps (3-timestep hit-probability shown; left purple field) could be faithfully reconstructed (**i**) using the same second half of the blue network PSI. This is particularly informative given that the agent never had access to information more than 1 timestep back, while the derived hit-probability is for 3 timesteps in the future: action values allow the agent to build up a longer-term predictive model.