A Neural Circuit Framework for Economic Choice: From Building Blocks of Valuation to Compositionality in Multitasking

Aldo Battista[1*], Camillo Padoa-Schioppa[2,3,4] and Xiao-Jing Wang[1*]

[1]Center for Neural Science, New York University, New York, NY, USA.
[2]Department of Neuroscience, Washington University in St. Louis, St. Louis, MO, USA.
[3]Department of Economics, Washington University in St. Louis, St. Louis, MO, USA.
[4]Department of Biomedical Engineering, Washington University in St. Louis, St. Louis, MO, USA.

*Corresponding author(s). E-mail(s): aldo.battista@nyu.edu; xjwang@nyu.edu;
Contributing authors: camillo@wustl.edu;

Abstract

Value-guided decisions are at the core of reinforcement learning and neuroeconomics, yet the basic computations they require remain poorly understood at the mechanistic level. For instance, how does the brain implement the multiplication of reward magnitude by probability to yield an expected value? Where within a neural circuit is the indifference point for comparing reward types encoded? How do learned values generalize to novel options? Here, we introduce a biologically plausible model that adheres to Dale's law and is trained on five choice tasks, offering potential answers to these questions. The model captures key neurophysiological observations from the orbitofrontal cortex of monkeys and generalizes to novel offer values. Using a single network model to solve diverse tasks, we identified compositional neural representations—quantified via task variance analysis and corroborated by curriculum learning. This work provides testable predictions that probe the neural basis of decision making and its disruption in neuropsychiatric disorders.

Introduction

Economic choice—the process by which individuals make decisions based on subjective preferences—is fundamental to both human and animal behavior [1, 2]. These decisions range from everyday choices to complex considerations involving multiple attributes like reward types (e.g. water versus grapefruit juice), quantity, probability, and delay. Neuro-scientific studies of economic choice at the single-cell level took off around the turn of this century [3–6]. Understanding economic decisions relies on the concept of *subjective value*, a measure that facilitates the comparison of different choices [6]. By assigning values to available options and making choices based on these values, the brain reduces complex, multidimensional decisions to a single dimension, facilitating efficient decision making [7].

The *orbitofrontal cortex* (OFC) has been identified as a key region supporting good-based decisions. Studies in non-human primates have revealed three groups of neurons in OFC essential for economic choice: *offer value neurons*, encoding the value of individual options; *chosen value neurons*, representing the value of the selected option; and *chosen good neurons*, indicating the identity of the chosen good [6]. These neurons exhibit *menu invariance*, maintaining consistent encoding regardless of alternative options—a property supporting choice transitivity [8]. Electrical stimulation studies have established a causal link between OFC neuronal activity and choice behavior in support of OFC's integral role in the decision circuit [9, 10].

However, the *circuit mechanisms* underlying *value computation* and *value comparison* remain largely unknown [10, 11]. Although the OFC has been closely associated with good-based decisions, it is still an open question whether value computations occur locally within the OFC or are computed in upstream regions and subsequently relayed to the OFC. Specifically, it is unclear how the brain derives the values of individual goods from multiple features and then compares these values to drive decision making. Moreover, existing studies often focus on single-neuron analyses and binary choice tasks, which do not fully capture neural population dynamics as well as the complexity of real-world decisions involving multiple options and attributes [7].

One way to gain insights into the decision mechanisms is to build a credible computational model that solves the task [12–16]. Previous computational models addressed aspects of economic decisions with some limitations. Built on a biologically based neural circuit model of decision-making [17, 18], Rustichini and Padoa-Schioppa [13] proposed a network that demonstrates the sufficiency of three distinct OFC neuron types in reproducing economic choice behavior. However, their model relies on strong circuit assumptions that may not fully reflect the neural heterogeneity observed experimentally. Specifically, it assumes that offer value neurons and chosen good neurons are exclusively excitatory, while chosen value neurons are exclusively inhibitory, and that all neuronal populations exhibit solely positive encoding—that is, their firing rates increase monotonically with the decision variable. In contrast, empirical studies reveal that OFC neurons can display both positive and negative tuning. Our model shows that the key decision variables can be robustly encoded by both excitatory and inhibitory neurons, each exhibiting diverse tuning properties. This suggests that the strict segregation of neuronal roles assumed by the previous model might not be necessary to account for the full spectrum of neural responses observed during economic decision-making [6].

On the other hand, Song, Yang, and Wang [15] used trained recurrent neural networks (RNNs) to assess whether OFC-like units emerge through learning, testing the necessity of these neurons in economic choice. Although this approach allows for complex task training, it employs Gated Recurrent Units (GRUs), which incorporate dynamic gating mechanisms and adjustable time constants that lack clear biological counterparts. Moreover, the network is divided into separate actor and critic modules,

further limiting its biological plausibility [9]. The lack of clear biological counterparts hinders the model's applicability to real neural circuits.

To bridge these gaps, we developed a biologically plausible computational model combining the strengths of the previous approaches while overcoming their limitations. Our model consists of a continuous-time recurrent neural network of the "vanilla" type (with no GRUs) that adheres to *Dale's law* with 80% excitatory (E) and 20% inhibitory (I) neurons and long-range excitatory projections [19, 20]. Neurons have biologically realistic single-unit time constants. We trained the network using the *Proximal Policy Optimization (PPO)* reinforcement learning algorithm [21, 22], balancing computational efficiency with the capacity to solve complex tasks.

We found that, after training, the network successfully performed a diverse array of economic choice tasks. Our model replicates key behavioral patterns observed experimentally, including choice consistency, risk attitudes, and order biases [7, 23]. This suggests that the model can serve as a platform for investigating the circuit mechanisms underlying choice biases [24], with implications for understanding neuropsychiatric disorders characterized by impaired decision making [1, 2].

Analysis of single-neuron activity within our model reveals cell groups mirroring those found in OFC: offer value neurons, chosen value neurons, and chosen good neurons. Notably, both excitatory and inhibitory neurons in our network are selective to decision variables, exhibiting *heterogeneous tuning* that aligns with experimental observations [6]. At the neural population level, we uncovered *low-dimensional dynamics* where specific directions in neural activity space correspond to decision variables [25].

We show how a multiplication of reward magnitude and probability is approximately computed for expected values in a neural network where synaptic input currents are additive for excitation and subtractive for inhibition. Furthermore, we found that the *relative values*—the essence of economic choice—are encoded in the *input weights* to the decision network where value computation occurs. Notably, this feedforward mechanism enhances the model's ability to generalize to *unseen offers*, addressing the critical generalization problem in real-world decision making [26]. This novel finding offers a compelling experimental prediction that synaptic efficacies play a crucial role in value computation.

Furthermore, our results suggest a novel mechanism for value computation occurring *upstream* of the decision circuit, while for value comparison, our model demonstrates that decisions are implemented via winner-takes-all (WTA) dynamics within the recurrent network [17, 18, 27]. This mechanism provides a computational framework for both binary and more complex choices, supporting the sufficiency of the identified neuron types in reproducing economic choice behavior without restrictive circuit assumptions [2].

Importantly, our findings highlight the *compositionality* of neural representations within the model. We show that a single neural circuit, with minimal variations, can solve multiple economic choice tasks. This is quantified by *task variance analysis* [28, 29], *rule subspace analysis*, and the use of *curriculum learning protocols* that accelerate training [30, 31].

In summary, our model generates several testable predictions for future experimental studies. It provides a biologically plausible model that bridges single-neuron observations and population-level dynamics and offers valuable insights into the functioning of OFC in economic decisions.

Results

Training neural networks for multiple economic choice tasks

To investigate the neural mechanisms underlying economic decisions, we developed biologically plausible excitatory-inhibitory recurrent neural networks (RNNs) capable of performing a range of complex economic choice tasks (Fig. 1**a**). These tasks were designed to capture key aspects of decision making, requiring the networks to compute and compare values in diverse contexts involving different goods, quantities, probabilities, and temporal structures.

Each task began with a fixation period, followed by a rule cue indicating the current task, the presentation of offers, and finally, a response phase. Choice tasks included the *standard task*, where two goods were offered in varying quantities; the *risky task*, similar to the standard task but with probabilistic outcomes; the *bundles task*, where offers consisted of bundles of two goods; the *ternary task*, involving choices among three goods; and the *sequential task*, where two goods were presented sequentially in random order, and choices relied on the network's working memory [7].

The networks were designed to be biologically plausible, consisting of continuous-time vanilla RNNs with excitatory and inhibitory neurons, obeying Dale's law (Fig. 1**b**) [19, 20, 32]. Inputs included fixation signals, quantities, probabilities of the offered goods, and task-specific rule cues. The network produced two distinct outputs: one readout provided the policy for action selection (analogous to the "actor"), and the other computed a value function (analogous to the "critic") that predicts the expected discounted future reward (or return). Specifically, the policy output determined the probabilities over available actions at each time during the trial, while the value function output estimated the expected return. This dual-readout architecture not only guided correct action selection but also demonstrated that the network could compute and evaluate the value of the presented options.

We trained the networks using Proximal Policy Optimization (PPO) [21], a reinforcement learning algorithm suitable for optimizing performance in complex tasks, which simultaneously optimizes both action selection and value estimation. This approach mirrors how animals are typically trained in laboratory tasks (i.e., through trials and errors, with reward feedback [22]) and is thus more biologically plausible than supervised learning. The training involved an agent-environment interaction loop, where the network received inputs and selected actions leading to new stimuli and action outcomes. Networks were trained separately on multitasking and curriculum learning protocols to study compositionality and learning-to-learn across different tasks [28, 30] (see Methods for details).

The networks achieved high performance across all tasks, satisfying the criteria that were set to reproduce the behavioral patterns observed in animals performing similar tasks: at least 99% of trials completed without fixation breaks and at least 90% correct choices among those trials (Fig. S1). Correct choices were defined as selecting the offer with the highest value among those presented. While it is possible to train networks to perform the tasks perfectly, our goal was to develop biologically plausible models to generate neurophysiological predictions, and thus, we aimed for performance levels comparable to those of animals in experimental settings [7].

To illustrate the network's behavior, we present sample trials for each choice task from a network trained simultaneously on all tasks (Fig. S2). In each trial, the network selects the highest-value offer among those presented. The value function output predicts the expected return shortly after the offers are presented, demonstrating the network's ability to compute the options' value. The policy outputs show the correct action selection during the response phase and indicate that the network maintains fixation during the required periods. Interestingly, the forthcoming choice can often be inferred during the offer presentation phase, even before the response period begins.

This observation suggests that the network undertakes two decision processes – an "economic" decision between two goods followed by a "motor" decision about when to reveal the choice outcome.

In the sequential task, we analyzed the example trial shown in Figure S2**e**. In this trial, the network receives the first offer and maintains in memory this information during the subsequent delay period. Upon presentation of the second offer, the network compares it with the first offer. During the response phase, the network selects the higher-value offer, indicating effective integration of sequential information. This performance suggests that the network has developed working memory capabilities, as it must retain information about the first offer over the delay period to make the correct choice in the sequential task. Without such working memory processes, the network would be unable to compare the offers and select the higher value.

These results demonstrate that our RNNs can effectively perform multiple and complex choice tasks. Indeed, the networks achieved high accuracy across all tasks and were able to process different types of information, such as quantities, probabilities, and temporal sequences, required for these tasks. Hence,the networks have developed the necessary computational mechanisms to perform value-based decisions in diverse contexts.

Behavioral patterns consistent with a value-based decision process

To assess whether 20 networks trained in a multitasking setting replicate decision processes observed in primates, we analyzed their choice behavior across all tasks. It is worth noticing that, like monkeys, the networks consistently select the offer with the highest computed value in each task.

Choice data were analyzed using logistic regression [7]. For example, in the risky task, each trial involves choices between goods C and E, each varying in quantity and probability. Initially, plotting the networks' choices in the space of quantities and probabilities does not reveal a clear decision boundary (Fig. 2**a**, left and center panels). However, when we transform the data into an *offer value space*—calculating the offer values as the product of the intrinsic value (ρ_X), quantity (q_X), and probability (p_X) raised to a power (γ)—a linear separation emerges (Fig. 2**a**, right panel and **b**). Here, γ quantifies the network's risk attitude, and ρ_X represents the inferred relative value between goods (see Methods).

Extending this analysis to the other tasks reveals a consistent strategy across the networks. Indeed, in the standard task, networks compare the computed values of two goods based on their quantities and intrinsic values, reliably selecting the good with the higher value. Similarly, in the bundles task, networks compute the total value of each bundle by summing the values of individual goods and choosing the bundle with the higher total value. In the ternary task, despite the increased complexity of comparing three options, networks reliably select the good with the highest computed value. In the sequential task, networks effectively maintain the value of the first offer in working memory and compare it with the second offer to make the optimal choice (Fig. S3).

Using a logistic regression model, we subsequently identified key behavioral parameters from the networks' choices across all tasks. Notably, the inferred *relative values* (ρ_X) match the intrinsic values assigned during training across all tasks (Fig. 2**c**). This consistency indicates that networks have learned the relative values of different goods and applied them in various contexts.

We also examined *behavioral biases* such as risk attitude and order bias. For tasks involving probabilistic outcomes, such as the risky, bundles, ternary, and sequential tasks, we estimated the risk attitude parameter (γ) for each network (Fig. 2**d**). A $\gamma > 1$ indicates risk aversion, while $\gamma < 1$ signifies risk-seeking behavior. On average, networks do not exhibit significant risk biases, aligning with the unbiased training objective.

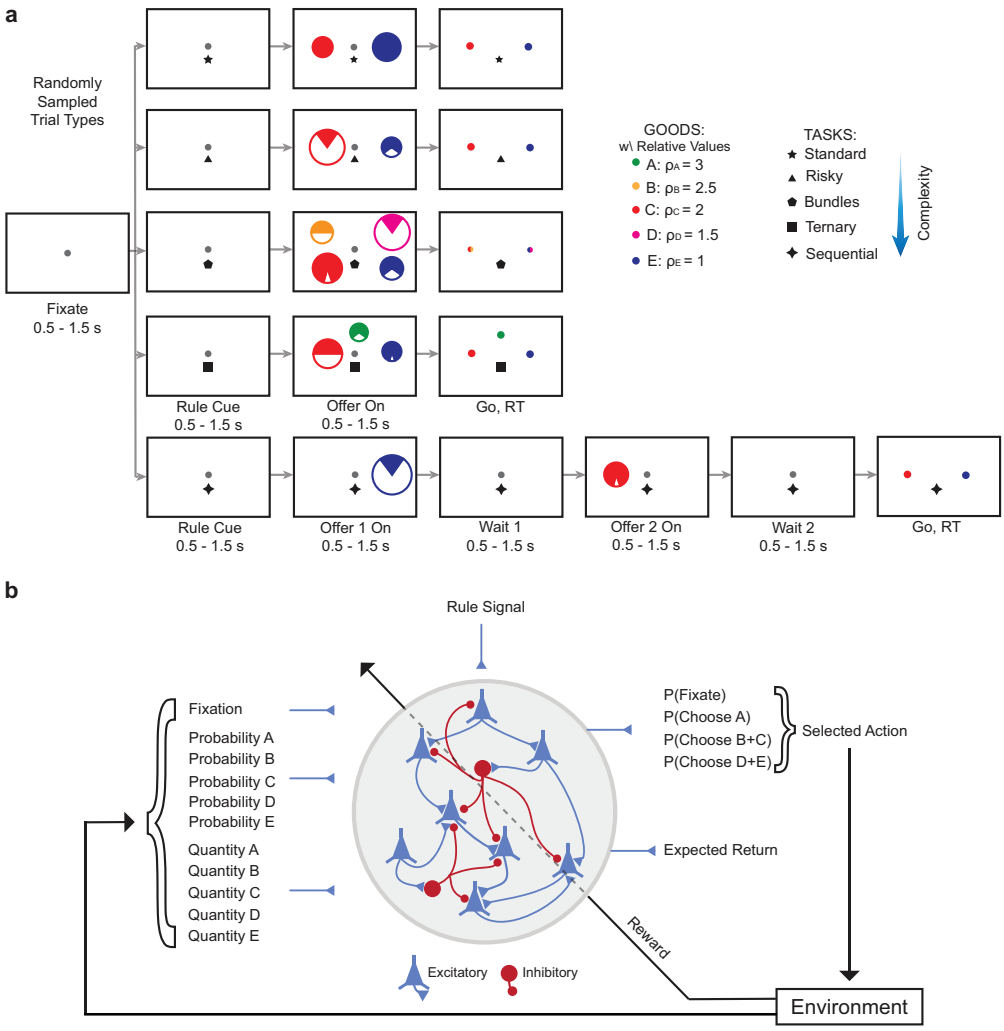

Fig. 1 Task structures and network architecture. a, Schematic representation of the economic choice tasks. Each task begins with a fixation period, followed by a rule cue indicating the current trial's task. Offers are then presented, followed by a response phase. Tasks include the standard task, risky task, bundles task, ternary task, and sequential task. Circles represent goods with varying intrinsic values (goods A–E), quantities (circle radius), and probabilities (filled area). Colors correspond to the intrinsic values of the goods. **b**, Structure of the biologically plausible recurrent neural network trained with reinforcement learning. The network consists of excitatory (E) and inhibitory (I) neurons adhering to Dale's law. Inputs include fixation signals, quantities, probabilities, and rule cues. Outputs include policy readouts for action selection and a value function predicting expected return. The network is trained using Proximal Policy Optimization in an agent-environment interaction loop.

However, individual networks display variability, with some showing mild risk-seeking or risk-averse tendencies. In the sequential task, we assessed order bias (ϵ'), which quantifies a preference for either the first or second offer regardless of their values. Again, while the average order bias across networks is negligible, individual networks may exhibit slight preferences, reflecting stochastic fluctuations during learning.

Finally, we assessed the relationship between *task complexity and choice accuracy*. The logistic regression provides a measure of choice consistency (η) proportional to the slope of the psychometric function. We found that networks perform better on simpler tasks like the standard task and exhibit lower accuracy on more complex tasks such as the sequential task (Fig. 2**e**). This pattern aligns with empirical observations in non-human primates, suggesting that task difficulty impacts decision performance [7].

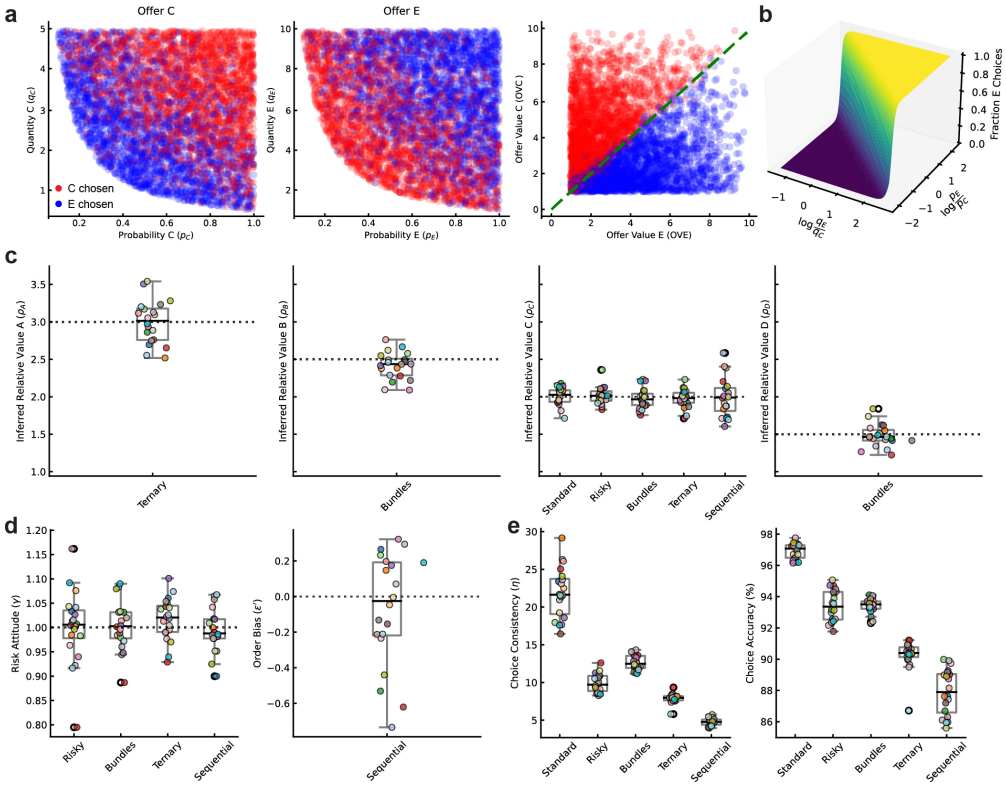

Fig. 2 Logistic behavioral analysis of 20 networks trained on all economic choice tasks. **a**, Logistic regression analysis for the *risky task* in a representative network. **Left and center panels**: Each point represents a trial, plotted in the space of quantities and probabilities for goods C and E, colored by the network's choice (red for C, blue for E). No clear decision boundary is apparent in this space. **Right panel**: The same trials plotted in *offer value space*, with offer values computed as $\mathrm{OV}_X = \rho_X \times q_X \times p_X^{\gamma}$. In this space, choices are linearly separable, indicating that the network bases decisions on computed offer values. **b**, Psychometric function obtained from logistic regression, showing the probability of choosing good E as a function of the logarithm of the ratio of offer features (quantities and probabilities). The sigmoid curve indicates consistent decision making based on offer values. **c**, Inferred relative values (ρ_X) for each good across different tasks, estimated from logistic regression on 20 networks (each network represented by a different color). The inferred values closely match the intrinsic values assigned during training, and the relative ranking of goods is preserved ($\rho_A > \rho_B > \rho_C > \rho_D > \rho_E$). **d**, Behavioral biases estimated from logistic regression. **Left**: Risk attitude parameter (γ) for tasks involving probabilistic outcomes. Values of $\gamma > 1$ indicate risk aversion, while $\gamma < 1$ indicate risk-seeking behavior. **Right**: Order bias (ϵ') in the sequential task, with positive values indicating a preference for the second offer. While biases are minimal on average, individual networks exhibit variability. **e**, Choice consistency (η) and choice accuracy across tasks for all networks. Choice consistency is proportional to the slope of the psychometric function from logistic regression, and choice accuracy reflects the network's ability to select the highest-value offer. Networks perform better on simpler tasks and show reduced performance on more complex tasks, such as the sequential task.

Single-neuron signatures of value computation and comparison

We assessed whether individual neurons in our networks represented decision variables similar to those observed in the orbitofrontal cortex (OFC) of non-human primates engaged in similar choice tasks [6, 33]. First, we focused on tasks involving binary choices between juices C and E (standard, risky, and sequential tasks). As in neurophysiology studies, we defined a series of candidate variables that could potentially explain the activity of individual cells, including individual offer values (OVC, OVE), the chosen value (CV), the other (non-chosen) value (OV), the chosen good (CG), the

value sum (OVC + OVE), the value difference (CV - OV). We also defined two good-specific chosen values (chosen value C, chosen value E), which represent the value of a good (C or E) when that good is chosen and zero otherwise.

The correlation matrix shown in Fig. 3**a** revealed significant patterns due to the task design. In particular, the chosen value correlates positively with the maximum of the offer values and with the value sum, as it represents the higher of the two offer values. Similarly, the choice variable is strongly correlated with the value difference between the offers since larger differences make the choice more deterministic. The conjunctive variables chosen value C and chosen value E are also correlated with their respective offer values and choices, reflecting their composite nature.

Analyzing the tuning properties of individual neurons, we found that many neurons displayed significant linear relationships with specific decision variables. Figure 3**b** illustrates examples of neurons tuned to different variables in the risky task. One neuron showed activity that increased linearly with the offer value of good E, another neuron's firing rate correlated with the chosen value, and a third neuron was selective for the binary choice of good E. These neurons exhibited significant coefficients of determination ($R^2 > 0.3$, $p < 0.05$), indicating robust tuning (see Methods).

To investigate the dynamics of the neuronal selectivity during a trial, we calculated the fraction of neurons selective for each decision variable at each time point (Fig. 3**c**; see Methods). During the offer presentation phase, we observed a peak in the fraction of neurons encoding offer values, reflecting the initial computation of individual offer values. As the trial progressed, the proportion of neurons encoding the chosen value increased, followed by an increase in the fraction of neurons encoding the choice. This temporal sequence mirrors the decision process where offer values are first computed and then compared to elaborate the final choice.

For the sequential task, by stimulus onset, we indicate the onset of the *second* stimulus. At this point, the network holds information about the first stimulus in working memory and processes the second stimulus, enabling it to compare both offers to make a decision (see Methods). This approach ensures that our analysis captures the period when the network has access to all relevant information for the decision.

Interestingly, both excitatory and inhibitory neurons exhibited similar dynamics, indicating that inhibitory neurons actively participate in encoding the decision variables [27], contrary to models that assign inhibitory neurons a non-selective role [13].

To assess the temporal stability of the neuronal tuning, we computed a Temporal Stability Index (TSI) for each neuron (Fig. 3**d**; see Methods). The majority of neurons showed high TSI values, indicating consistent encoding of a single decision variable throughout the trial. However, some neurons displayed lower TSI scores, suggesting dynamic coding where neurons might switch from encoding offer value to encoding choice as the decision process unfolds. This dynamic tuning aligns with observations in OFC, where neurons can change their selectivity over the course of a decision [33].

We further investigated whether neurons encode decision variables categorically or conjunctively by analyzing the distribution of differences in R^2 values between pairs of decision variables, focusing on neurons that were selective (i.e., had $R^2 \geq 0.3$ for at least one of the variables) (Fig. 3**e**; see Methods) [34]. Bimodal distributions for the pairs of offer value versus chosen value and chosen value versus choice suggest that neurons tend to encode one variable over the other, supporting categorical encoding. This specialization allows for a more distinct representation of decision variables within the network.

Examining the sign of encoding, we found that both excitatory and inhibitory neurons displayed a mix of positive and negative correlations (encodings) with their respective decision variables (Fig. 3**f**; see Methods). This suggests that neurons can either increase or decrease their firing rates with increasing values of the decision variable—a phenomenon consistent with empirical findings in OFC [6]. This heterogeneity

in tuning enhances the network's capacity to represent information and indicates that inhibitory neurons play an active role in processing decision-related signals [27].

In conclusion, the single-neuron analyses reproduce key features of neuronal encoding observed in OFC of non-human primates during economic choice tasks [6]. The sequential activation of neurons encoding offer value, chosen value, and choice reflects the computational stages of decision making. The active participation of inhibitory neurons and the presence of both positive and negative tuning expand our understanding of the neural mechanisms underlying value computation and comparison. These findings support the notion that economic decisions emerge from distributed computations within recurrent neural circuits, with neurons dynamically encoding relevant variables to guide behavior.

Population-level signatures of value computation and comparison

To further understand how value computation and comparison are implemented in the trained networks, we analyzed the population dynamics of recurrent neurons during economic choice tasks. While single-neuron analyses revealed specific neurons encoding decision variables, population-level analyses can uncover how these variables are represented collectively and how dimensionality reduction techniques can capture the main components of the decision process.

We first examined the dimensionality of the recurrent neural activity across different tasks during the stimulus presentation. Using principal component analysis (PCA), we quantified the number of dimensions required to explain approximately 85% of the variance in the population activity, using the participation ratio as a measure of embedding dimensionality (Fig. 4**a**; see Methods) [35]. Interestingly, we found that for all tasks except the ternary task, the neural activity was low dimensional (two-dimensional), suggesting that the population dynamics are constrained to low-dimensional manifolds, likely reflecting the encoding of key decision variables such as offer value, chosen value, and choice. In the ternary task, which involves choices among three goods, a third dimension was necessary to capture the additional complexity of the decision space. Moreover, when excitatory and inhibitory neurons were separated, we observed that both populations exhibited similar dimensionality patterns, indicating that inhibitory neurons contribute actively to the encoding of decision variables at the population level. The slightly higher participation ratio for inhibitory neurons may be partially attributed to their smaller population size (only 20% of the network).

To illustrate the population dynamics during a specific task, we first focused on the risky task in a representative network. Projecting the neural activity onto the first two principal components, we found that the first principal component (PC1) encoded the chosen value, while the second principal component (PC2) encoded the choice (Fig. 4**b**). Each point in the plot represents a trial, and the separation along these components reflects how the network differentiates between decision variables at the population level. This pattern was consistent across both excitatory and inhibitory neurons. To visualize how these dynamics evolve over time, we projected the population activity onto the first two principal components throughout time (Supplementary Video 1). It reveals that as time progresses, the neural trajectories corresponding to different choices and offer values diverge along the principal components, highlighting the temporal unfolding of value computation and comparison in the network. We subsequently extended this analysis across all tasks and networks. By performing a linear regression of the projections onto the principal components against various decision variables, we quantified how much variance each component explained for each variable (Fig. 4**c**; see Methods). The results showed that, except for the ternary task, the first principal component consistently encoded the chosen value, while the second component encoded the choice. In the ternary task, the additional dimension captured by the

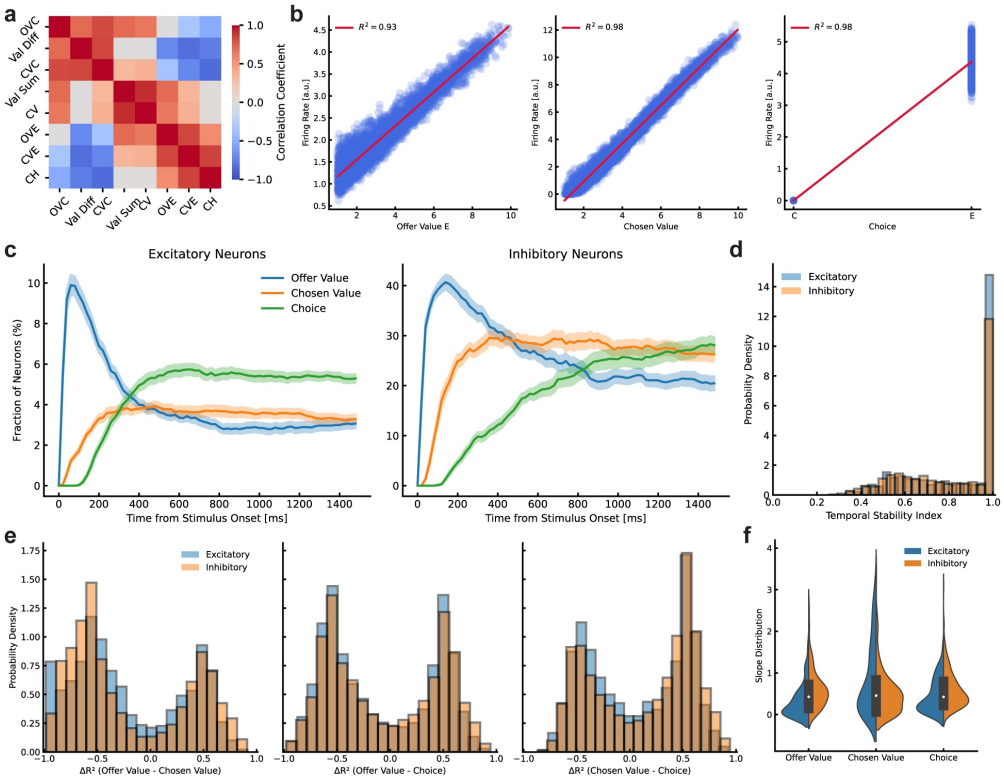

Fig. 3 Single-neuron analysis in networks trained on all economic choice tasks. a, Correlation matrix between behavioral variables for tasks involving choices between two goods (standard, risky, and sequential tasks). The matrix shows intrinsic correlations among variables such as offer values (OVC, OVE), chosen value (CV), conjunctive variables (chosen value C, chosen value E), value sum, value difference, and choice (CH), averaged across twenty networks. Notably, chosen value correlates with value sum, and choice correlates with value difference. **b**, Examples of tuning curves from individual neurons in a trained network for the risky task. Left: Neuron encoding the offer value of good E (OVE). Middle: Neuron encoding chosen value (CV). Right: Neuron encoding choice of good E (CH). Each point represents a trial, plotting the mean firing rate against the respective decision variable. Red lines indicate linear regression fits with corresponding R^2 values. **c**, Fraction of neurons selective for each decision variable over time during the stimulus presentation phase, averaged across tasks and networks. For the sequential task, time zero corresponds to the onset of the second stimulus. Left: Excitatory neurons. Right: Inhibitory neurons. Shaded areas represent the standard error of the mean. **d**, Temporal Stability Index (TSI) distributions for excitatory and inhibitory neurons. TSI measures the consistency of a neuron's selectivity for the selected variable (the variable encoded for the majority of time steps) over time. **e**, Categorical encoding analysis. Histograms of differences in R^2 values between pairs of decision variables for selective neurons (with $R^2 \geq 0.3$ for at least one variable), separately for excitatory and inhibitory neurons. Bimodal distributions suggest categorical encoding. **f**, Distribution of regression slopes for neurons selective to each decision variable, showing both positive and negative encoding among excitatory and inhibitory neurons.

third principal component was necessary to encode the choices among the three goods. At the population level, excitatory and inhibitory neurons showed similar encoding patterns, reinforcing the notion that inhibitory neurons play an active role in processing decision-related signals.

To better investigate the role of recurrent connectivity in shaping population dynamics, we performed a lesion analysis by removing all recurrent connections from the trained networks (see Methods). This manipulation effectively eliminated the recurrent dynamics while preserving the feedforward inputs to the network. After lesioning, we observed significant changes in the distribution of firing rates during the stimulus presentation phase. Specifically, the mean firing rates of both excitatory and inhibitory neurons increased compared to the original networks (Supplementary Fig. S4). This

increase is likely due to the loss of inhibitory feedback and recurrent competition that typically regulates neuronal activity levels. We repeated the dimensionality analysis on the lesioned networks (as done on the original networks). Interestingly, the dimensionality increased for most tasks, suggesting that the recurrent connections contribute to constraining the neural activity into lower-dimensional manifolds (Supplementary Fig. S5**a**). In the sequential task, however, the dimensionality decreased, reflecting the network's inability to maintain information of the first offer in working memory without recurrent connections. In the lesioned networks, PCA analysis revealed that the first principal component primarily encoded the value sum, while the second component encoded the value difference (Supplementary Fig. S5**b**), that is a linear combination of the offer values. In particular, chosen value and choice variables were not encoded; without recurrent dynamics, the networks could no longer perform the decision process.

This orientation of the principal component axes—where PC1 reflects the sum of the offer values rather than the individual offer values—suggests that in the absence of recurrent dynamics, the network's feedforward processing organizes the representation to emphasize the combined (or total) value. This structure likely emerges as an adaptation to facilitate a downstream readout process, whereby the chosen value (which is strongly correlated with the value sum) can be more readily extracted, even though the raw inputs still contain the individual offer values.

Moreover, we observed that the neural trajectories in the lesioned network during the risky task remain clustered according to offer values but do not exhibit the separation seen in the intact networks (Supplementary Video 2). This suggests that while the offer values are computed upstream from the inputs, the recurrent network dynamics are crucial for comparing these values and generating a decision.

In summary, these population-level analyses highlight the critical role of upstream computations in value computation and recurrent connectivity in implementing value comparison through WTA dynamics.

Dissecting the circuit mechanisms of value computation

To uncover how the networks compute offer values, we first examined the input weights and their contributions to the multiplicative computations required for choices between probabilistic outcomes (e.g., in the risky task). By analyzing the input weight matrices after training, we observed a distinct structure indicating that value computation occurs upstream of the recurrent decision circuit. In particular, the correlation matrix of input weight vectors for quantities and probabilities revealed a block structure in which input weights associated with the same offer were strongly positively correlated, while those corresponding to different offers were uncorrelated (Fig. 5**a**). This suggests that the recurrent neurons are organized into subpopulations specialized for processing each offer, integrating quantity and probability features to compute offer values. The segregation of input weights points to a mechanism wherein the networks learn to approximate the multiplication of quantity and probability for each offer, be it a single good or a bundle of goods.

Next, to determine whether these networks are indeed computing the offer values through multiplication rather than addition, we consider the case of lesioned recurrent connections again and focus on feedforward computations. We projected population activity onto the first two principal components and performed linear regressions against both the product and the sum of quantities and probabilities (Fig. 5**b**). The resulting coefficients of determination (R^2) were significantly higher for the product than for the sum, confirming that the networks approximate multiplicative computations. Both excitatory and inhibitory neurons contributed to this process, with the principal components reflecting the offer values of different goods.

To further understand how weighted sums and nonlinear activations might approximate multiplication, we built a simplified feedforward network model. Input weights

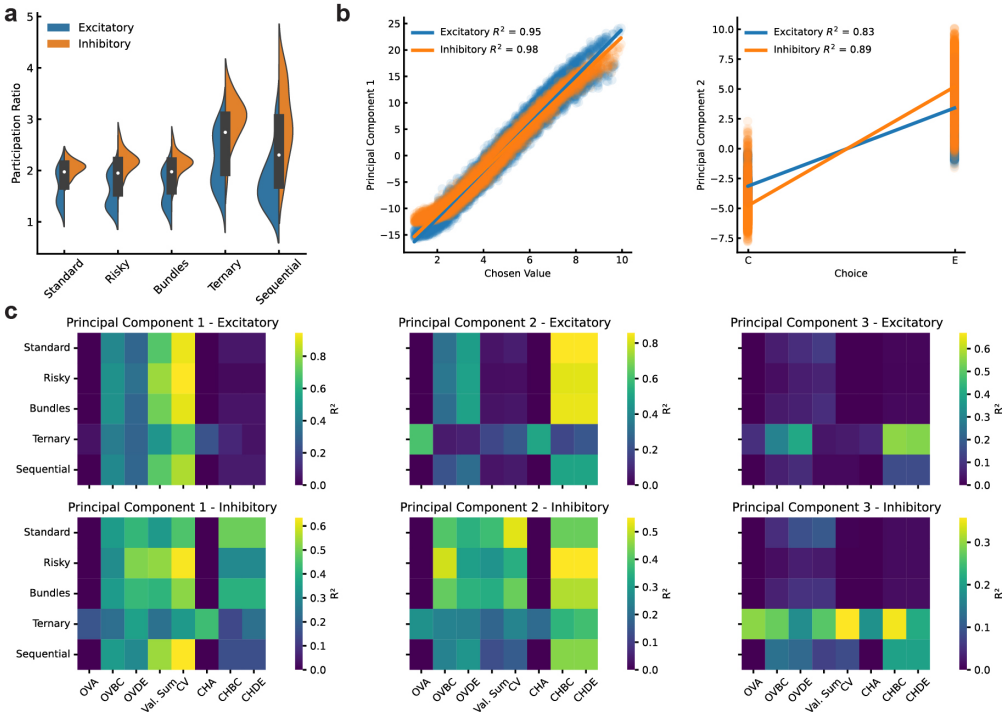

Fig. 4 Population analysis of networks trained on all economic choice tasks. a, Neural dimensionality across different networks and tasks, measured by the participation ratio, which estimates the number of dimensions required to explain approximately 85% of the variance in population activity [35] during the stimulus presentation phase. Analyses were conducted separately for excitatory (blue) and inhibitory (orange) neurons. All tasks are predominantly two-dimensional except for the ternary task, which requires a third dimension, suggesting low-dimensional dynamics associated with key decision variables. **b**, Principal component analysis (PCA) of neural activity during the risky task in a representative network. Scatter plots of the projections onto the first two principal components (PC1 and PC2) are shown separately for excitatory and inhibitory neurons. Each point represents a trial. Linear regression indicates that PC1 primarily encodes the chosen value, while PC2 encodes the choice. For a dynamic visualization, see Supplementary Video 1. **c**, Summary of population analyses across all networks and tasks. Heatmaps display the average coefficients of determination (R^2) from linear regression of the principal components onto various decision variables. Rows represent different tasks, columns represent decision variables, and the color intensity reflects the R^2 value. The analyses confirm that PC1 and PC2 encode chosen value and choice-related variables, respectively, across tasks, with similar patterns observed for excitatory and inhibitory neurons. PC3 is mainly involved in the ternary and sequential tasks that require additional computations (e.g., value computation of the third good and working memory, respectively).

scale linearly with neuron indices, and neurons are divided into two groups processing features of different goods. In this toy model, neurons receive weighted inputs of quantity and probability, pass them through a rectified linear unit (ReLU) activation function, and produce outputs that approximate the product of the inputs at the population level (Supplementary Fig.S6; see Methods). Principal component analysis of the hidden units' activity showed that the first two principal components corresponded to the offer values of the two goods (Fig.5**c**). However, this model alone did not explain the rotation of the offer value axes observed in the trained networks. In the intact networks, the feedforward computations initially align the principal component axes with the individual offer values. Yet, when recurrent dynamics are removed—as in our lesion analysis—the representation rotates so that PC1 predominantly captures the sum of the offer values and PC2 reflects their difference. This rotated coordinate system suggests that the network is reconfiguring its representation to better extract a single, unified decision variable (the chosen value), which is highly correlated with the value

sum. In essence, the rotation is not an inherent property of the multiplicative computation itself but rather an emergent consequence of the network's need to read out the chosen value from combined offer information. We, therefore, extended the toy model by introducing a population of linearly mixed-selective neurons that receive inputs from both goods and compute linear combinations of their features, ultimately encoding the value sum (which is correlated with the chosen value). The introduction of this mixed-selectivity population rotates the offer value axes relative to the principal components and aligns the model more closely with our empirical observations (Fig. 5**d**). This rotation facilitates the computation of the chosen value through subsequent recurrent dynamics.

These results, revealing that offer values are computed upstream of the recurrent circuit, support the conceptualization of economic choice as a two-stage process of value computation and value comparison.

Next, we examined how the networks implement the relative value between goods, which captures the quintessence of economic choice behavior. We hypothesized that the relative values learned during training are embedded within the input weights that connect specific goods' features to the recurrent neurons. To test this, we trained 50 networks exclusively on the risky task, systematically varying the intrinsic value of one good. After training, we used logistic regression to infer the relative values from choice behavior and then correlated these with the average non-zero input weights for the quantity inputs of the high-value good (Fig. 5**e**). We observed a strong linear correlation, confirming that these relative values are indeed encoded in the input weights. This finding suggests that by adjusting only the input weights, the network could readily generalize to new goods with different intrinsic values without necessitating any alteration to the recurrent circuitry responsible for value comparison.

These analyses indicate that value computation occurs upstream of the recurrent decision circuit, with offer values being computed through multiplicative integration of features in the input layer. The recurrent network then implements value comparison via WTA dynamics. This modular organization implies that the decision circuit is robust to changes in goods and their values, requiring only adjustments in input weights to accommodate choices between novel goods. Such a mechanism supports flexible decision making and may reflect how biological neural circuits generalize across different contexts and experiences.

Dissecting the circuit mechanisms of value comparison

Having established that value computation occurs upstream of the recurrent decision circuit, we next investigated how the network compares these computed values to make decisions. We focused on the dynamics of the network's outputs and the underlying recurrent connectivity patterns that facilitate value comparison through winner-take-all (WTA) dynamics.

We first analyzed the activity of the network's output units during the risky task, averaging across all trials (splitting by choice) and networks. The outputs corresponding to choices of goods C and E exhibited WTA dynamics (Fig. 6**a**). When good C was chosen, the output unit associated with C showed increased activity during the stimulus presentation phase, while the output for E remained suppressed, and vice versa. This mutual inhibition between choice outputs indicates a competitive process where the representation of the higher-valued option dominates, leading to a decision.

Interestingly, the fixation output maintained the highest activity during the stimulus presentation, reflecting the network's requirement to sustain fixation until the response phase. In particular, linear regression analysis revealed that the fixation output also encoded the chosen value (Fig.6**b**; see Methods). This suggests that the network integrates information about the expected reward into the fixation output to modulate decision timing, preventing early choices before the fixation cue turns off. The

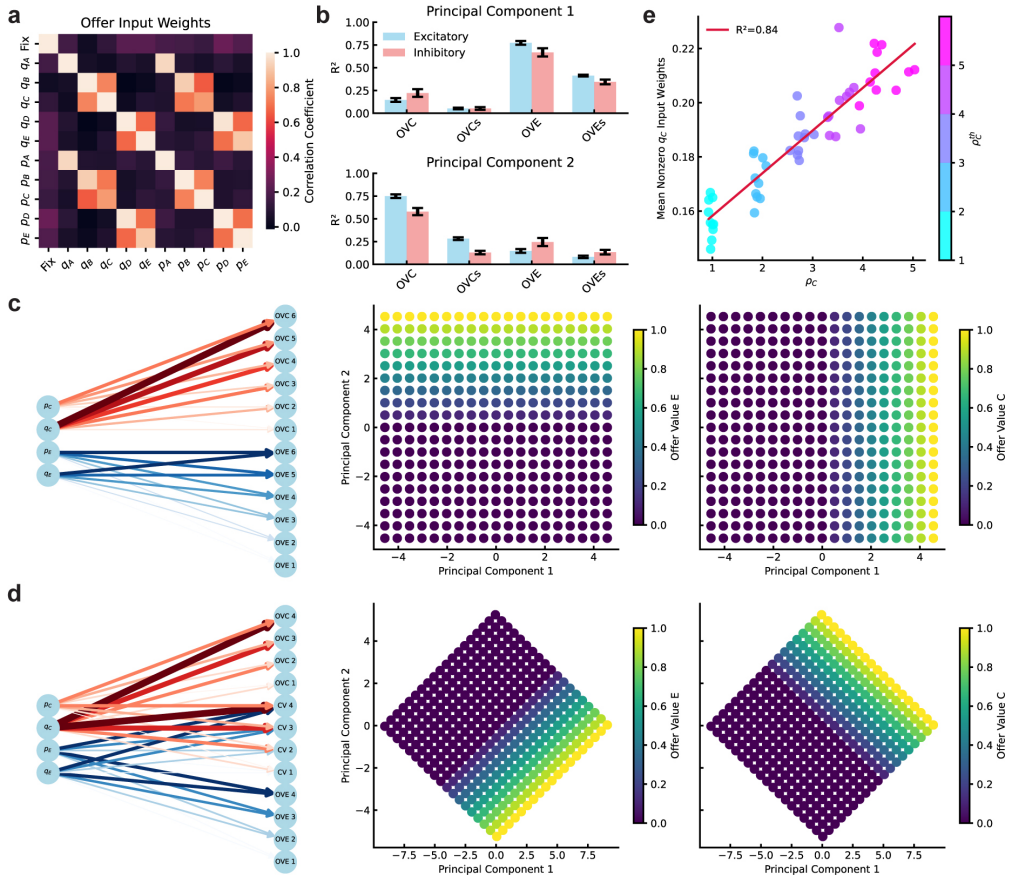

Fig. 5 Dissecting the circuit mechanisms of value computation. **a**, Correlation matrix of input weight vectors for quantities and probabilities associated with different goods in a trained network. Strong positive correlations are observed between input weights corresponding to the same offer, indicating a block structure where recurrent neurons are organized into subpopulations processing specific offer features. **b**, Comparison of multiplicative (OVC and OVE) and additive (OVC_s and OVE_s) models in the risky task. Regression analyses show higher coefficients of determination (R^2) when projecting population activity onto the first two principal components and regressing against the product (multiplication) of quantity and probability compared to the sum. Both excitatory (blue) and inhibitory (orange) neurons contribute to this computation. **c**, Toy feedforward network model approximating multiplication through weighted sums and nonlinear activation functions. Input weights scale linearly with neuron indices, and neurons are divided into two groups processing features of different goods. PCA of the hidden units' activity reveals axes corresponding to the offer values of the two goods. **d**, Extension of the toy model including a third population of mixed-selectivity neurons receiving inputs from both goods. This results in a rotation of the offer value axes relative to the principal components, consistent with observations in the trained networks. **e**, Correlation between the relative values inferred from logistic regression of choice behavior and the average non-zero input weights for the quantity input of the high-value good across networks trained with varying intrinsic values. The strong linear relationship indicates that relative values are encoded in the input weights.

dynamics of the outputs show two distinct aspects of the decision process: determining which option to choose and deciding when to choose it. The fact that the forthcoming choice can be inferred from the output activity before the response phase implies that reaction times (RTs), measured as the difference between the time of the choice and the onset of the response phase, are not significantly influenced by task difficulty or value differences between options. This was confirmed by analyzing RT distributions across tasks, which showed similar RTs regardless of the absolute value difference, indicating that the decision is made before the action is executed (Supplementary Fig.S7**a**).

To understand the neural mechanisms underlying these WTA dynamics, we examined the recurrent connectivity matrices of the trained networks. Singular value decomposition (SVD) revealed that the recurrent weight matrices were low-rank, with only a few singular values needed to explain most of the variance (Supplementary Fig. S7**b**). This low-rank structure suggests that the network's dynamics are governed by a limited number of connectivity patterns or motifs, which in turn constrain the neural activity to low-dimensional subspaces that efficiently encode the critical decision variables. This low-dimensional organization reduces noise and focuses variability along task-relevant dimensions, thereby enhancing the efficiency of decision readout.

We further dissected the network by categorizing neurons based on their selectivity for decision variables—offer value, chosen value, and choice—and their excitatory or inhibitory nature. By averaging the recurrent weights between these neuron groups and across different trained networks, we constructed a reduced connectivity matrix (Fig. 6**c**; see Methods). This matrix revealed specific motifs consistent with competitive dynamics. For instance, excitatory neurons selective for a particular choice had strong positive connections to inhibitory neurons selective for the same choice. The inhibitory neurons, in turn, provided negative feedback to excitatory neurons encoding the opposing choice, thereby reinforcing the selection of the higher-valued option.

Visualizing these connectivity patterns in a simplified circuit diagram (Fig. 6**d**), we observed that excitatory choice-selective neurons not only promote their own activity via self-excitation but also inhibit competing choices through inhibitory interneurons. This reciprocal inhibition mediates the WTA dynamics necessary for value comparison. Additionally, inhibitory neurons selective for the chosen value interact with excitatory choice neurons, further shaping the decision process. In our network, the feedforward inputs and activation functions generally produce uniformly positive responses. However, the selective inhibitory interactions can effectively invert these responses. Specifically, when an inhibitory neuron tuned to a particular decision variable suppresses an excitatory neuron with similar tuning, the net output of that excitatory neuron may exhibit a negative relationship with the decision variable—its firing rate decreases as the variable increases. This targeted inhibitory feedback, by inverting the response slope, generates a heterogeneous tuning profile wherein some neurons encode decision variables with a positive slope while others encode them with a negative slope. Such an arrangement is critical because it sharpens the contrast between competing options, ensuring that one option distinctly "wins" (attaining high activity) while the other is actively suppressed (attaining low activity).

We term this connectivity-mediated process the Competitive Recurrent Inhibition (CRI) mechanism. Unlike a mere phenomenon of WTA—where one option simply wins, and the other loses—the CRI mechanism specifies how structured recurrent interactions (including both excitatory and inhibitory connections) actively mediate this outcome. Notably, our results reveal that the CRI mechanism, while reminiscent of the recurrent competition described in Wang (2002) [17], exhibits distinct connectivity motifs that are responsible for both the amplification of the chosen option and the inversion of neuronal tuning, thereby enabling robust value comparison across diverse tasks.

We extended this analysis to other tasks and found similar connectivity motifs in the reduced connectivity matrices (Supplementary Fig. S7**c**). This consistency across tasks supports the idea that the CRI mechanism is a general strategy employed by the network to implement value comparison, ensuring that in every decision context one option emerges with high activity and the alternatives are suppressed to low activity levels. Future lesion experiments targeting these specific connectivity motifs (e.g., selectively removing inhibitory feedback from choice-selective neurons) would further clarify which connections are necessary and sufficient for proper decision-making.

In summary, our findings demonstrate that the recurrent network performs value comparison through the CRI mechanism—structured recurrent connectivity that not

only implements winner-take-all dynamics but also generates heterogeneous neuronal tuning (both positive and negative encodings) to maximize contrast between competing offers. The low-rank nature of the recurrent connectivity supports low-dimensional dynamics that efficiently encode task-relevant information, providing insight into how neural circuits integrate value information and make decisions in line with processes observed in biological systems.

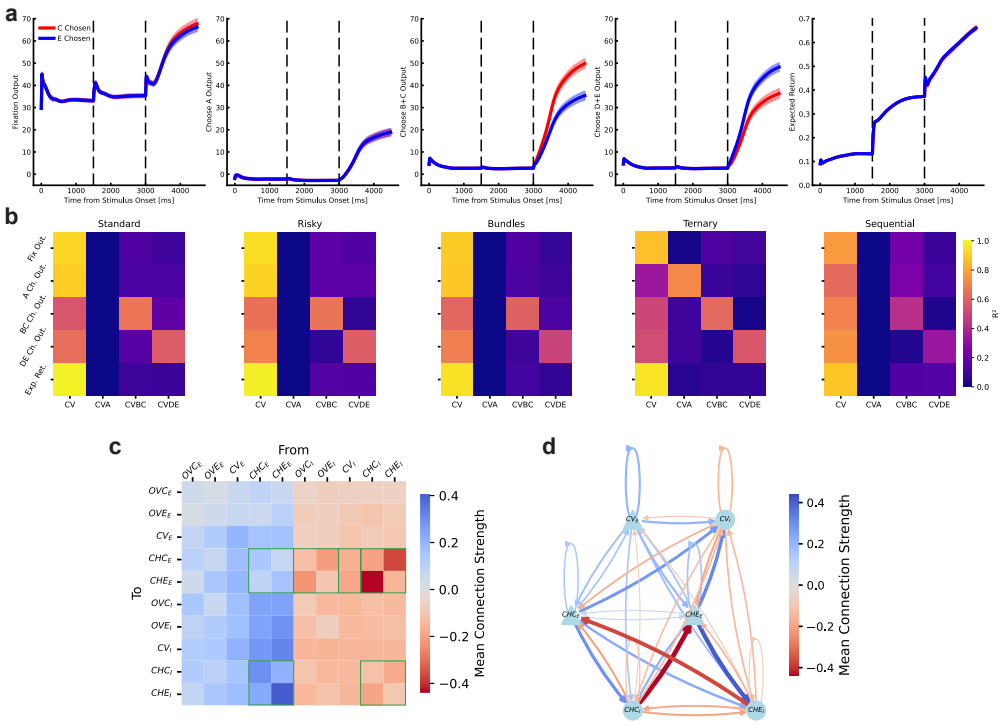

Fig. 6 Dissecting the neural circuit mechanisms of value comparison. a, Average activity of the network's output units during the risky task, separated by trials where good C (red) or good E (blue) was chosen. Data are averaged across all trained networks, with error bars representing the standard error of the mean. The outputs corresponding to choices of goods C and E exhibit winner-take-all dynamics, with the chosen good's output dominating during the stimulus presentation phase. The fixation output remains high throughout, reflecting the need to maintain fixation until the response phase. **b**, Linear regression analysis of the output units' activity during the last 200 ms before the response phase against decision variables. The fixation output encodes both fixation and the chosen value, while the expected return output encodes the chosen value by design. Choice outputs primarily encode the value of the corresponding good when that good is chosen, reflecting the winner-take-all dynamics. **c**, Reduced recurrent connectivity matrix for the risky task, showing the average weights between neuron populations categorized by selectivity for decision variables and neuronal type (excitatory or inhibitory). The matrix reveals motifs consistent with competitive recurrent inhibition (CRI), highlighted by green rectangles. **d**, Simplified circuit diagram focusing on the choice and chosen value populations. The diagram illustrates the competitive interactions between excitatory and inhibitory neurons encoding different choices, facilitating the selection of the higher-valued option through recurrent dynamics.

Compositionality in economic decisions

To understand how the networks handle multiple economic choice tasks, we investigated the extent to which neural representations are shared across tasks versus specialized for specific tasks, focusing on schema formation and compositionality. By analyzing networks trained on all tasks, we aimed to discern how neural circuits flexibly adapt to different task demands while reusing common computational mechanisms.

We began by examining the rule input weight vectors—*i.e.*, the synaptic weights connecting the scalar rule cue (which signals to the network the specific task to be solved on a given trial) to each of the N recurrent neurons. Correlation analysis of these N-dimensional vectors across the five tasks revealed that the rule input weight vectors for all tasks except the sequential task were highly aligned, indicating a shared rule representation. In contrast, the sequential task exhibited a distinct rule input weight vector, reflecting its unique working memory demands (Fig. 7**a**). This suggests that the network utilizes a shared input structure for most tasks, making the explicit rule cue irrelevant. In contrast, the sequential task exhibited distinct input weights for its rule cue, reflecting its unique requirement for working memory to maintain information about sequentially presented offers.

Further analysis of the neural representations underlying each task showed that the population dynamics during the rule-cue period occupied similar subspaces for all tasks except the sequential task. By calculating the participation ratio to estimate the dimensionality of the neural activity and computing the angles between the subspaces spanned by different tasks, we found that the neural subspaces for the standard, risky, bundles, and ternary tasks were highly overlapping (Fig. 7**b**; see Methods). The sequential task, however, occupied a distinct subspace, consistent with its reliance on working memory processes not required by the other tasks. These findings indicate that the network engages similar neural dynamics for most tasks, utilizing a shared circuit for task resolution, while the sequential task recruits additional neural resources due to its unique computational demands.

We subsequently analyzed the variance of neuronal firing rates during the stimulus presentation phase across tasks to delve deeper into how individual neurons contribute to different tasks. By clustering neurons based on their normalized variance profiles, we identified distinct groups (Fig. 7**c**). One cluster consisted of neurons active across all tasks, representing shared computational components such as value computation and fixation maintenance. Another cluster was specific to the ternary task, likely involved in processing the additional good unique to that task. A third cluster was specific to the sequential task, reflecting neurons engaged in working memory processes.

We quantified the relationships between tasks by examining the distributions of task variance differences for each neuron across task pairs (Fig. 7**d**; see Methods). The histograms revealed patterns of inclusive relationships, where one task's neural representation is a subset of another's, and disjoint relationships, where tasks engage distinct neuronal populations. These results support the idea of a flexible network architecture that balances common processing with task-specific adaptations, enabling efficient and compositional decision making.

To visualize the compositionality across tasks, we projected the mean firing rates of all neurons during stimulus presentation into the space of the first three principal components. Plotting these projections for all networks showed that the points corresponding to different tasks clustered together, with the exception of the sequential task, which formed a separate cluster (Fig. 7**e**). This separation underscores the unique neural dynamics required by the sequential task due to its working memory demands, highlighting the network's ability to form specialized representations when necessary.

Finally, we investigated how prior learning influences the acquisition of new tasks using curriculum learning protocols. Networks trained sequentially on simpler tasks before progressing to more complex ones exhibited accelerated learning compared to networks trained from scratch (Fig. 7**f**). This suggests that the networks formed a schema—a shared set of computational strategies—that facilitated knowledge transfer across tasks. The reuse of learned components, such as temporal processing and multiplicative value computations, enabled the networks to efficiently adapt to new task demands, reflecting the brain's ability to leverage prior experience for improved learning and performance in novel situations.

In summary, our analyses demonstrate that networks trained on multiple economic choice tasks develop compositional architectures that combine shared circuitry with task-specific adaptations. The formation of common schemas and the ability to reuse computational modules support flexible and efficient decision making across diverse contexts.

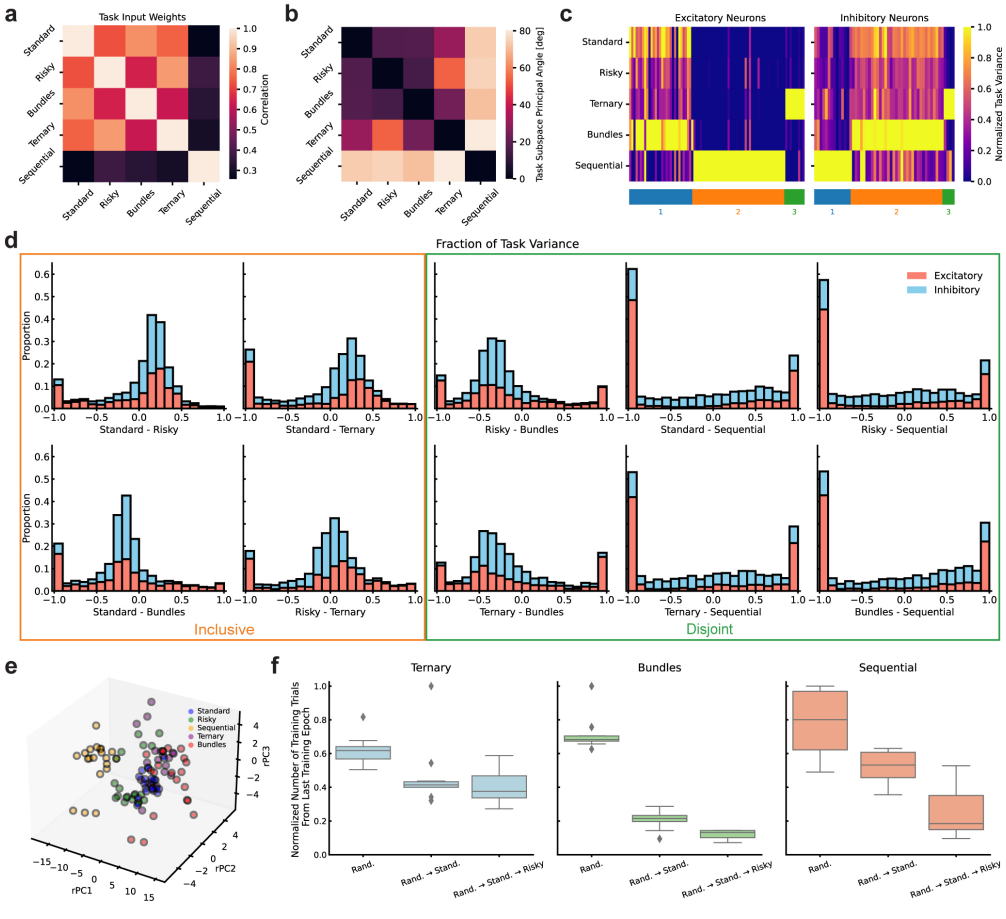

Fig. 7 Compositionality and curriculum learning in networks trained on all economic choice tasks. a, Correlation matrix of input weight vectors for rule cues across tasks, averaged over all networks. High correlations are observed among all tasks except the sequential task, indicating shared input structures. The sequential task has distinct input weights due to its unique working memory requirements. **b**, Subspace analysis of population activity during the rule cue period. The participation ratio estimates the dimensionality, and the angles between subspaces of different tasks reveal that all tasks except the sequential task occupy overlapping subspaces, suggesting shared neural representations. **c**, Clustering of neurons based on task variance of firing rates during stimulus presentation. Heatmaps show normalized variance for each neuron (columns) across tasks (rows), with neurons grouped by k-means clustering. Clusters include neurons active across all tasks (shared components), neurons specific to the ternary task, and neurons specific to the sequential task (specialized components). **d**, Histograms of neurons task variance differences for task pairs, illustrating inclusive and disjoint relationships between tasks. **e**, Visualization of task representations by projecting mean firing rates onto the first three rotated principal components (aligned between different networks). Points corresponding to different tasks cluster together, except for the sequential task, which forms a separate cluster, highlighting its distinct neural dynamics. **f**, Curriculum learning analysis showing accelerated learning in networks trained sequentially on tasks. Learning curves compare networks trained from scratch to those using curriculum learning, demonstrating the benefits of schema formation and reuse of computational components across tasks.

Circuit mechanisms for generalization in value computation

Understanding how neural circuits generalize learned value computations to novel situations is crucial for explaining flexible decision making in dynamic environments. Recent studies have highlighted the brain's capacity for generalization in value-based choices, where subjects apply learned valuation strategies to stimuli outside their prior experience [26]. To explore the neural mechanisms underlying such generalization, we examined how our networks trained on the risky task handle offers that were never encountered during training.

We trained ten networks only on the risky task using a constrained set of offers. In one condition, probabilities were fixed at half of their maximum possible value while quantities varied across the full available range; in the other condition, quantities were fixed at half the maximum while probabilities varied fully (Fig. 8**a**; see Methods). This training regime ensured that the networks learned to compute offer values based on either varying quantities or probabilities, but not both simultaneously.

After training, the networks successfully chose the higher-valued offer within the constrained training set, indicating effective learning of value computation under limited conditions. To assess generalization, we then tested the networks on an unconstrained set where both quantities and probabilities varied across their full ranges, presenting offers that the networks had not seen during training (Fig. 8**b**). The networks demonstrated robust generalization, accurately selecting the higher-valued offer despite the novel combinations of quantities and probabilities.

Behavioral analysis using logistic regression confirmed that the inferred relative values and risk attitudes remained consistent between the constrained training set and the unconstrained test set (Fig. 8**c**; see Methods). While choice consistency and accuracy were slightly lower in the test set, the overall performance remained high, suggesting that the networks effectively generalized their value computations.

This generalization arises from the networks' computation of the multiplication between quantity and probability when computing the offer values. Even though the networks were trained on limited combinations of these variables, their approximation of multiplication allowed them to interpolate and extrapolate to novel offer values. The network's ability to generalize is thus rooted in their computation of offer value as the product of quantity and probability, enabling them to apply learned valuation strategies to new situations.

Our findings align with observations in non-human primate studies, where subjects generalize valuation processes to novel stimuli [26]. The networks' generalization demonstrates how neural circuits can leverage fundamental computational principles, such as approximate multiplication, to extend learned behaviors beyond specific training experiences.

Discussion

In this work, we developed a biologically plausible computational model of recurrent neural networks (RNNs) with excitatory-inhibitory neurons for economic decisions. The salient findings are fivefold. First, the model reproduces salient single-neuron and population activity patterns observed in the OFC of monkeys performing economic choice tasks. In particular, inhibitory cells are as selective as excitatory cells, in consonance with observations of single-neuron activity and synaptic connectivity [36, 37]. Second, we identified a circuit mechanism for computing the expected values approximately as a product of reward probability and quantity. This multiplication computation explains the network's ability to generalize to values of unseen novel choice options, as observed experimentally [26]. Third, the relative value of different goods, which imposes specific indifference points, is encoded in the input weights. Fourth, our network analyses provide direct insight into the circuit mechanisms underlying winner-take-all (WTA)

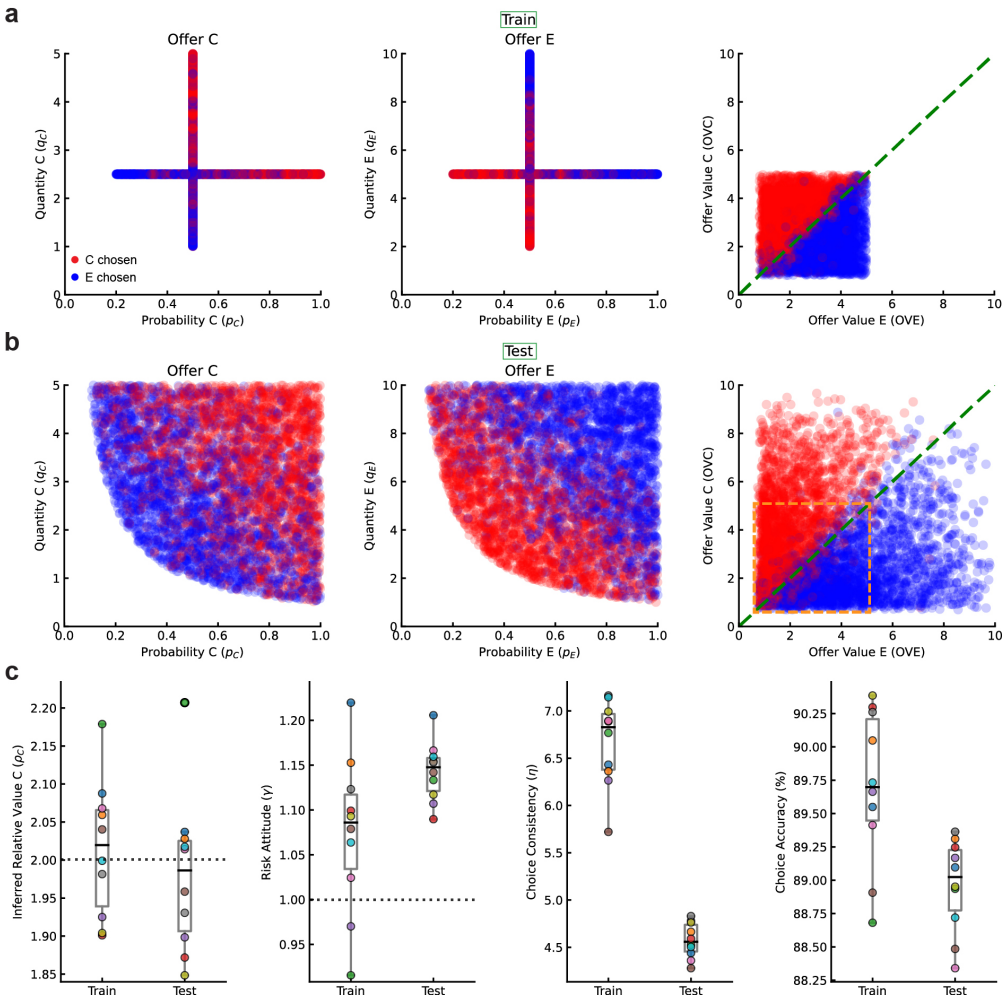

Fig. 8 Generalization in 10 **networks trained on a constrained risky task. a**, Training regime where networks were trained on a constrained set of offers in the risky task. In one condition, probability was fixed at half of its maximum value while quantity varied across the full range. In the other condition, quantity was fixed at half maximum while probability varied fully. **b**, Testing the networks on an unconstrained set where both quantity and probability varied across their full ranges. Networks demonstrated good generalization, accurately selecting the higher-valued offer in novel situations. **c**, Summary of behavioral analysis for the ten trained networks. Points represent individual networks, showing that inferred relative values and risk attitudes remained consistent between the constrained training set (Train) and the unconstrained test set (Test). Choice consistency and accuracy were slightly lower in the test set but remained high overall, indicating effective generalization.

dynamics: distinct patterns of recurrent connectivity between excitatory and inhibitory neurons orchestrate competitive interactions that selectively amplify the representation of the higher-valued option while suppressing lower-value signals, thereby implementing value comparison. Fifth, neural representations display compositionality, which accelerates learning when training to perform multiple tasks is temporally organized according to an appropriate curriculum. These findings offer novel insights into how economic choices are represented and processed in the brain, which are useful for future investigations.

Although direct anatomical evidence for fixation signals, explicit task-rule cues, and probability inputs to the OFC is still debated, our model uses these signals as abstract representations of the sensory and contextual information that the decision

circuitry receives. This assumption not only provides a plausible framework for integrating external cues but also generates testable predictions regarding the role of these inputs in value computation.

A significant achievement of our model is its ability to replicate key behavioral patterns observed in empirical studies, including choice consistency, risk attitudes, and order biases [7, 23]. These consistencies demonstrate the model's robustness and its utility as a tool for exploring the neural substrates of choice biases [24], with implications for understanding neuropsychiatric disorders characterized by impaired decision making, such as frontotemporal dementia, schizophrenia, and drug addiction. The observed variability in risk attitude and order bias across networks is likely due to stochastic fluctuations during training—such as random initializations and input noise—that can break symmetry and predispose a network toward risk-seeking or risk-averse behavior. Future work will focus on systematically varying these noise sources to elucidate their specific contributions to bias formation.

At the behavioral level, human and animal choices are typically "as if" based on the computation of subjective values [38]. Supporting this construct, neurons in OFC and other brain regions explicitly represent the values of offered and chosen goods. However, it is, in principle, possible to make effective choices without computing values [39–43]. Thus a critical aspect of our results is that our networks did indeed compute and compare values. More specifically, our analysis revealed that neurons in the model exhibit tuning properties similar to those observed in OFC, with different cells representing variables offer value, chosen value, and chosen good [1, 6]. At the neuronal level, our analysis revealed that neurons in the model exhibit tuning properties that closely mirror those observed in the OFC. In particular, distinct neurons encode key decision variables—offer value, chosen value, and chosen good. Importantly, both excitatory and inhibitory neurons display heterogeneous tuning, meaning that not only do all these neurons encode the relevant task variables, but they also do so with diverse response profiles. Specifically, some neurons show a positive monotonic relationship with the encoded variable (i.e., their firing rate increases as the variable increases), while others exhibit a negative monotonic relationship (i.e., their firing rate decreases as the variable increases). This diversity in tuning, spanning both cell types and both directionalities of response, aligns with experimental observations and challenges previous models that assumed a more uniform, exclusively positive encoding scheme. While negative encoding emerges naturally from the structured inhibitory feedback in our model, we propose that these negative responses are not merely epiphenomenal. Instead, they may actively contribute to enhancing the contrast between competing offers, thereby improving the efficiency of value comparison. This prediction remains to be tested in future experimental studies. The sequential dynamics of these neurons—first computing value and then comparing values—further reflect the temporal structure of decision making observed in OFC [13]. The dynamics of these neurons follow the sequential process of value computation followed by value comparison, reflecting the temporal structure of decision making [2, 6].

At the population level, our model demonstrates low-dimensional dynamics where specific neural activity patterns correspond to decision variables [25]. This low-dimensional structure suggests that economic decisions are driven by a set of variables capturing most of the variance in neural activity, with chosen values potentially encoded as line attractors [44]. The discovery that relative values are encoded in the input synaptic weights offers a compelling experimental prediction, suggesting that synaptic efficacies play a crucial role in value computation.

Our model proposes a novel mechanism for value computation occurring upstream of the decision circuit, approximating the multiplication of goods' features such as probability and quantity [11, 45, 46]. This mechanism enhances the model's ability

to generalize to unseen offers, addressing a critical aspect of real-world decision making [26]. The capacity for generalization suggests that the brain may employ similar computational strategies to navigate complex environments efficiently. It is important to note that our model defines expected reward as a multiplicative function of quantity and probability. However, alternative formulations—such as an additive combination—might capture behavior under different conditions, particularly under varying levels of uncertainty. Recent work [47] suggests that the optimal strategy may depend on context. Future work should explore how these alternative computations influence decision performance.

For value comparison, our model demonstrates that decisions are implemented via WTA dynamics within the recurrent network [17, 18, 27]. This mechanism provides a computational framework for both binary and complex choices, supporting the sufficiency of the identified neuron types in reproducing economic choice behavior. The connectivity structure observed in our networks provides testable predictions for future experiments *in vivo*.

Moreover, the identification of both specialized and shared neural clusters across different economic choice tasks suggests that the brain efficiently reuses neural circuits for multiple tasks while maintaining the ability to specialize for specific decision contexts when necessary. The emergence of dedicated neural mechanisms for working memory functions in sequential tasks highlights the flexibility and adaptability of neural circuits in supporting complex cognitive functions [32, 48, 49].

Our study also highlights the significance of compositionality and multitasking in neural circuits. The model's ability to perform multiple tasks by recombining simpler, previously learned components reflects the inherent flexibility of neural representations in the brain [28, 29]. This compositionality, coupled with curriculum learning protocols [30, 31], accelerates the learning process and allows adaptation to diverse economic contexts. These findings suggest that the brain may employ similar strategies to optimize learning and decision making, balancing specialization and generalization [50, 51].

Future work is needed to test the proposed neural circuit model, with neural data recorded from non-human primates and rodents [6, 52, 53]. Such efforts will validate the model's predictions and refine our understanding of the neural circuits involved in economic choice. Testing our predictions—such as the role of inhibitory neurons in encoding decision variables and the emergence of negatively encoding cells due to selective inhibition (for example, in our model, inhibitory neurons that are selectively tuned for a specific decision variable can target and suppress other neurons with similar tuning. This selective inhibition can invert the effective tuning slope—from a positive, feedforward-driven response to a negative one—since, without inhibitory interactions, all encoding would be exclusively positive by construction. Such an emergent inversion, driven by the winner-take-all dynamics in the recurrent network, constitutes a key prediction of our framework that can be empirically tested)—using advanced neurophysiological techniques will further confirm the biological plausibility of our framework [9, 10].

The model's ability to generalize across various decision contexts prompts further investigation into the neural mechanisms supporting this flexibility. Understanding how neural circuits navigate the trade-off between specialization and generalization could provide deeper insights into the principles governing economic decisions [51, 54].

Our model can be extended and improved in multiple ways. First, a model can be modified to receive and process the actual input images the animals see on laboratory screens instead of simplified scalar input representations where the different features are disentangled for simplicity [55]. Second, such modeling could be used for comparison between the OFC and other brain regions involved in decision making, such as the ventromedial prefrontal cortex and the amygdala [56–59]. Third, a frontier topic is to

investigate valuation [60] and decision-making [61] widely distributed in a large-scale brain system underlying economic choices.

References

1. Padoa-Schioppa, C. Neurobiology of economic choice: a good-based model. *Annual review of neuroscience* **34,** 333–359 (2011).
2. Padoa-Schioppa, C. & Conen, K. E. Orbitofrontal cortex: a neural circuit for economic decisions. *Neuron* **96,** 736–754 (2017).
3. M. L. Platt & P. W. Glimcher. Neural correlates of decision variables in parietal cortex. *Nature.* **400,** 233–238 (1999).
4. L. P. Sugrue, G. C. Corrado & W. T. Newsome. Matching behavior and representation of value in parietal cortex. *Science* **304,** 1782–1787 (2004).
5. D. J. Barraclough, M. L. Conroy & D. Lee. Prefrontal cortex and decision making in a mixed-strategy game. *Nat. Neuroscience* **7,** 404–410 (2004).
6. Padoa-Schioppa, C. & Assad, J. A. Neurons in the orbitofrontal cortex encode economic value. *Nature* **441,** 223–226 (2006).
7. Padoa-Schioppa, C. Logistic analysis of choice data: A primer. *Neuron* **110,** 1615–1630 (2022).
8. Padoa-Schioppa, C. & Assad, J. A. The representation of economic value in the orbitofrontal cortex is invariant for changes of menu. *Nature neuroscience* **11,** 95–102 (2008).
9. Ballesta, S., Shi, W., Conen, K. E. & Padoa-Schioppa, C. Values encoded in orbitofrontal cortex are causally related to economic choices. *Nature* **588,** 450–453 (2020).
10. Ballesta, S., Shi, W. & Padoa-Schioppa, C. Orbitofrontal cortex contributes to the comparison of values underlying economic choices. *Nature communications* **13,** 4405 (2022).
11. Raghuraman, A. P. & Padoa-Schioppa, C. Integration of multiple determinants in the neuronal computation of economic values. *Journal of Neuroscience* **34,** 11583–11603 (2014).
12. Solway, A. & Botvinick, M. M. Goal-directed decision making as probabilistic inference: a computational framework and potential neural correlates. *Psychological review* **119,** 120 (2012).
13. Rustichini, A. & Padoa-Schioppa, C. A neuro-computational model of economic decisions. *Journal of neurophysiology* **114,** 1382–1398 (2015).
14. Friedrich, J. & Lengyel, M. Goal-directed decision making with spiking neurons. *Journal of Neuroscience* **36,** 1529–1546 (2016).
15. Song, H. F., Yang, G. R. & Wang, X. Reward-based training of recurrent neural networks for cognitive and value-based tasks. *Elife* **6,** e21492 (2017).
16. Zhang, Z., Cheng, Z., Lin, Z., Nie, C. & Yang, T. A neural network model for the orbitofrontal cortex and task space acquisition during reinforcement learning. *PLOS Computational Biology* **14,** e1005925 (2018).
17. Wang, X.-J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36,** 955–968 (2002).
18. Wong, K.-F. & Wang, X.-J. A recurrent network mechanism of time integration in perceptual decisions. *Journal of Neuroscience* **26,** 1314–1328 (2006).
19. Song, H. F., Yang, G. R. & Wang, X. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLoS computational biology* **12,** e1004792 (2016).
20. Yang, G. R. & Wang, X. Artificial neural networks for neuroscientists: a primer. *Neuron* **107,** 1048–1070 (2020).

21. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
22. Sutton, R. & Barto, A. *Reinforcement learning: An introduction* (MIT press, 2018).
23. Shi, W., Ballesta, S. & Padoa-Schioppa, C. Economic choices under simultaneous or sequential offers rely on the same neural circuit. *Journal of Neuroscience* **42,** 33–43 (2022).
24. Fascianelli, V. *et al.* Neural representational geometries reflect behavioral differences in monkeys and recurrent neural networks. *Nature Communications* **15,** 6479 (2024).
25. Dubreuil, A., Valente, A., Beiran, M., Mastrogiuseppe, F. & Ostojic, S. The role of population structure in computations through neural dynamics. *Nature neuroscience* **25,** 783–794 (2022).
26. Bongioanni, A. *et al.* Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature* **591,** 270–274 (2021).
27. Roach, J. P., Churchland, A. K. & Engel, T. A. Choice selective inhibition drives stability and competition in decision circuits. *Nature communications* **14,** 147 (2023).
28. Yang, G. R., Joglekar, M. R., Song, H., Newsome, W. T. & Wang, X. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience* **22,** 297–306 (2019).
29. Driscoll, L. N., Shenoy, K. & Sussillo, D. Flexible multitask computation in recurrent networks utilizes shared dynamical motifs. *Nature Neuroscience,* 1–15 (2024).
30. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. *Curriculum learning* in *Proceedings of the 26th annual international conference on machine learning* (2009), 41–48.
31. Goudar, V., Peysakhovich, B., Freedman, D. J., Buffalo, E. A. & Wang, X.-J. Schema formation in a neural population subspace underlies learning-to-learn in flexible sensorimotor problem-solving. *Nature Neuroscience* **26,** 879–890 (2023).
32. Liu, Y. & Wang, X.-J. Flexible gating between subspaces in a neural network model of internally guided task switching. *Nature Communications* **15,** 6497 (2024).
33. Padoa-Schioppa, C. Neuronal origins of choice variability in economic decisions. *Neuron* **80,** 1322–1336 (2013).
34. Onken, A., Xie, J., Panzeri, S. & Padoa-Schioppa, C. Categorical encoding of decision variables in orbitofrontal cortex. *PLoS computational biology* **15,** e1006667 (2019).
35. Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv,* 214262 (2017).
36. Najafi, F. *et al.* Excitatory and inhibitory subnetworks are equally selective during decision-making and emerge simultaneously during learning. *Neuron* **105,** 165–179 (2020).
37. Kuan, A. T. *et al.* Synaptic wiring motifs in posterior parietal cortex support decision-making. *Nature* **627,** 367–373 (2024).
38. Kreps, D. M. *A course in microeconomic theory* (Princeton university press, 2020).
39. Tversky, A. & Simonson, I. Context-dependent preferences. *Management science* **39,** 1179–1189 (1993).
40. Gigerenzer, G. & Todd, P. M. *Simple heuristics that make us smart* (Oxford University Press, USA, 1999).
41. Lichtenstein, S. & Slovic, P. The construction of preference: An overview. *The construction of preference* **1,** 1–40 (2006).

42. Piantadosi, S. T. & Hayden, B. Y. Utility-free heuristic models of two-option choice can mimic predictions of utility-stage models under many conditions. *Frontiers in neuroscience* **9,** 105 (2015).
43. Soltani, A., De Martino, B. & Camerer, C. A range-normalization model of context-dependent choice: a new model and evidence. *PLoS computational biology* **8,** e1002607 (2012).
44. Pereira-Obilinovic, U., Hou, H., Svoboda, K. & Wang, X.-J. Brain mechanism of foraging: Reward-dependent synaptic plasticity versus neural integration of values. *Proceedings of the National Academy of Sciences* **121,** e2318521121 (2024).
45. Murphy, B. K. & Miller, K. D. Multiplicative gain changes are induced by excitation or inhibition alone. *Journal of neuroscience* **23,** 10040–10051 (2003).
46. Groschner, L. N., Malis, J. G., Zuidinga, B. & Borst, A. A biophysical account of multiplication by a single neuron. *Nature* **603,** 119–123 (2022).
47. Farashahi, S., Donahue, C. H., Hayden, B. Y., Lee, D. & Soltani, A. Flexible combination of reward information across primates. *Nature human behaviour* **3,** 1215–1224 (2019).
48. Mante, V., Sussillo, D., Shenoy, K. V. & Newsome, W. T. Context-dependent computation by recurrent dynamics in prefrontal cortex. *nature* **503,** 78–84 (2013).
49. Wang, J. X. *et al.* Prefrontal cortex as a meta-reinforcement learning system. *Nature neuroscience* **21,** 860–868 (2018).
50. Johnston, W. J. & Fusi, S. Abstract representations emerge naturally in neural networks trained to perform multiple tasks. *Nature Communications* **14,** 1040 (2023).
51. Johnston, W. J., Fine, J. M., Yoo, S. B. M., Ebitz, R. B. & Hayden, B. Y. Semi-orthogonal subspaces for value mediate a binding and generalization trade-off. *Nature Neuroscience,* 1–13 (2024).
52. Zhang, M. *et al.* The Representation of Decision Variables in Orbitofrontal Cortex is Longitudinally Stable. *bioRxiv,* 2024–02 (2024).
53. Kuwabara, M., Kang, N., Holy, T. E. & Padoa-Schioppa, C. Neural mechanisms of economic choices in mice. *Elife* **9,** e49669 (2020).
54. Veselic, S. *et al.* A cognitive map for value-guided choice in ventromedial prefrontal cortex. *bioRxiv* (2023).
55. Soo, W. W., Battista, A., Radmard, P. & Wang, X.-J. *Recurrent neural network dynamical systems for biological vision* in *The Thirty-eighth Annual Conference on Neural Information Processing Systems* ().
56. Hunt, L. T. *et al.* Mechanisms underlying cortical activity during value-guided choice. *Nature neuroscience* **15,** 470–476 (2012).
57. Rudebeck, P. H., Saunders, R. C., Lundgren, D. A. & Murray, E. A. Specialized representations of value in the orbital and ventrolateral prefrontal cortex: desirability versus availability of outcomes. *Neuron* **95,** 1208–1220 (2017).
58. Xia, F. *et al.* Understanding the neural code of stress to control anhedonia. *Nature,* 1–9 (2024).
59. Wang, X.-J., Jiang, J. & Pereira-Obilinovic, U. Bifurcation in space: emergence of function modularity in the neocortex. *bioRxiv* (2023).
60. Vickery, T. J., Chun, M. M. & Lee, D. Ubiquity and specificity of reinforcement signals throughout the human brain. *Neuron* **72,** 166–177 (2011).
61. Wang, X.-J. Theory of the multiregional neocortex: large-scale neural dynamics and distributed cognition. *Ann. Rev. Neurosci.* **45,** 533–560 (2022).

Acknowledgements

We thank G. Tavoni, C. Constantinople, E. Rich, and V. Fascianelli for their critical comments on the manuscript, as well as current and former members of the Wang lab, especially W. Soo, G. Benigno, Maryada, Y. Liu, V. Goudar, and A. Fanthomme, for fruitful discussions. This work was supported by the Swartz Foundation (A.B.); ONR grant N00014-23-1-2040 (X.-J.W.); NIH grants R01-MH104494, R01-DA032758, and R01-DA055709 (to C.P.-S.); and the NYU IT High Performance Computing resources, services, and staff expertise.

Author contributions

A.B., C.P.-S., and X.-J.W. conceptualized the study. A.B. trained the networks and analyzed the data in interactions with X.-J.W.. A.B., C.P.-S., and X.-J.W. wrote the manuscript.

Competing interests

The authors declare no competing interests.

Methods

Network structure

We employed excitatory-inhibitory (E/I) continuous-time vanilla recurrent neural networks (RNNs) [19, 28], adhering to key biological constraints to model the neural mechanisms underlying economic decision making (Fig. 1**b**). The networks consisted of $N = 256$ neurons, with 80% excitatory and 20% inhibitory neurons, reflecting cortical neuron ratios [62]. Neurons had a time constant $\tau = 100$ ms, consistent with cortical neurons with NMDA receptor-mediated synaptic dynamics [17]. The dynamics of the network were governed by:

$$\tau \frac{d\mathbf{r}(t)}{dt} = -\mathbf{r}(t) + \left[W^{\text{rec}}\, \mathbf{r}(t) + W^{\text{in}}\, \mathbf{u}(t) + \mathbf{b} + \sqrt{2\tau}\, \sigma_{\text{rec}}\, \boldsymbol{\xi}_{\text{rec}}(t) \right]_{+}, \tag{1}$$

where $\mathbf{r}(t)$ is the vector of neuronal activities, W^{rec} is the recurrent weight matrix, W^{in} is the input weight matrix, $\mathbf{u}(t)$ is the input vector, $\mathbf{b}$ is a bias term, $\boldsymbol{\xi}_{\text{rec}}(t)$ represents Gaussian white noise with zero mean and unit variance affecting the recurrent units, $\sigma_{\text{rec}} = 0.15$ is the standard deviation of the recurrent noise, and $[\cdot]_+$ denotes the rectified linear unit (ReLU) activation function. The network was simulated with a temporal discretization of $\delta t = 20$ ms.

The input vector $\mathbf{u}(t)$ included scalar representations of fixation, quantities and probabilities of the offered goods, and task-specific rule cues. All scalar inputs were normalized between 0 and 1, with an added baseline of $u_0 = 0.2$ and additive input noise:

$$\mathbf{u}(t) = \mathbf{u}_{\text{signal}}(t) + u_0 + \sqrt{2\tau}\, \sigma_{\text{in}}\, \boldsymbol{\xi}_{\text{in}}(t), \tag{2}$$

where $\mathbf{u}_{\text{signal}}(t)$ represents the normalized input signals, $\boldsymbol{\xi}_{\text{in}}(t)$ is Gaussian white noise with zero mean and unit variance affecting the inputs, and $\sigma_{\text{in}} = 0.01$ is the standard deviation of the input noise.

In our model, the fixation signal is provided as an input that mimics the experimental requirement for subjects to maintain gaze. It serves as a gating mechanism that holds the network in a non-committal state until the response phase, ensuring that value computations are performed while fixation is maintained.

The network architecture adhered to Dale's law by constraining excitatory neurons to have positive outgoing weights and inhibitory neurons to have negative outgoing weights, enforced by fixed masks during training [19]. Similarly, long-range projections (inputs and outputs) were enforced to be excitatory through masks applied during training, reflecting the excitatory nature of projections between brain areas [63].

Initialization of parameters. All weight matrices and biases were initialized as follows:

- **Recurrent weight matrix** W^{rec} was initialized by sampling each element from a Gamma distribution. Specifically, for each pair of neurons (i, j), the weight W^{rec}_{ij} was sampled from a Gamma distribution with shape and scale parameters equal to 4. After initialization, the weight matrix was rescaled to have a spectral radius (the largest absolute eigenvalue) of 1.5. A fixed mask was then applied to enforce Dale's law: weights corresponding to excitatory neurons (rows) were set to positive if negative, and weights corresponding to inhibitory neurons were set to negative if positive.

- **Input weight matrix** W^{in} was initialized by sampling each element from a uniform distribution over $[0, 1/\sqrt{N_{in}}]$, where $N_{in} = 16$ is the number of scalar inputs to the recurrent units. The inputs have excitatory projections to the recurrent units, and a mask was applied to enforce this constraint.

- **Bias terms b** were initialized to zero.

- **Output weights** $W^{\text{out}}_{\text{actor}}$ and $W^{\text{out}}_{\text{critic}}$ were initialized by sampling each element from a uniform distribution over $[0, 0.4/\sqrt{N}]$. The weights were kept non-negative, consistent with the excitatory nature of long-range projections. Masks were applied during training to maintain this constraint.

- **Output biases** $\mathbf{b}_{\text{actor}}$ and b_{critic} were initialized to zero.

The actor and critic shared the same recurrent network weights but had separate readouts. This design choice aligns with experimental observations that stimulation of value-related units can bias decisions [9]. Having separate networks for actor and critic, as in some architectures [15], would not be biologically plausible in this context.

Noise parameters. The standard deviation of the recurrent noise was set to $\sigma_{\text{rec}} = 0.15$, and the standard deviation of the input noise was set to $\sigma_{\text{in}} = 0.01$. These values were chosen to introduce variability in the neuronal activities and inputs, simulating the stochastic nature of neural processing in biological systems.

Activation function. The ReLU activation function was used to model the non-linear response of neurons, ensuring that neuronal activities remain non-negative, while unit-specific biases were introduced to model different activation thresholds for different units.

We ensured that the networks started from a plausible physiological state by specifying the initialization of all parameters, including weights and biases, and enforcing biological constraints through masking and proper scaling. The use of Gamma distribution for initializing the recurrent weights, followed by rescaling to a spectral radius of 1.5, promotes the emergence of dynamic activity patterns while maintaining stability [64].

Choice tasks and performance

We trained networks on a diverse set of economic choice tasks designed to simulate various aspects of decision making observed in primate studies [7]. These tasks required the networks to evaluate and compare offers involving different goods, quantities, probabilities, and temporal sequences. Each task presented unique challenges while sharing a common structure, necessitating different computational strategies.

We trained networks in three modes: (a) a multitask setting where a single network is trained concurrently on all five economic choice tasks; (b) single-task training, where separate networks are trained on individual tasks; and (c) a curriculum

learning approach in which the same network is sequentially trained on individual tasks—starting with simpler tasks and progressively introducing more complex ones. In the multitask networks, the same recurrent circuitry is used for all tasks, with task-specific differences arising primarily from the rule cue and associated input weights.

Common Task Structure. Each trial began with a fixation period of variable duration (500–1500 ms), during which the networks had to maintain fixation to proceed. This was followed by the presentation of a rule cue (500–1500 ms), indicating the specific task to be performed. The offer presentation phase varied depending on the task, and finally, a response period of up to 1000 ms was provided for the networks to make a choice. The duration of each phase was randomized to prevent reliance on temporal cues and favor the networks to learn fixed points (Fig. 1**a**).

Visual Representation and Network Inputs. In the task schematics, offers were visually represented by colored circles corresponding to different goods (A through E), each with a predefined intrinsic value: $\rho_A = 3$, $\rho_B = 2.5$, $\rho_C = 2$, $\rho_D = 1.5$, and $\rho_E = 1$. The circle's radius was proportional to the quantity offered, and in tasks involving probability, the filled area represented probability. This mimicked how such information might be presented to animals in experimental settings [7].

However, the networks did not process visual images. Instead, they received scalar inputs as proxies for these visual features. The networks received dedicated input channels for each good presented in a trial, encoding the quantity and probability (when applicable). Each good had its own input units for quantity and probability, allowing the networks to distinguish between different goods based on these inputs. The fixation cue and task rule cue were also provided as distinct scalar inputs. All inputs were normalized between 0 and 1, with an added baseline of $u_0 = 0.2$ and additive noise to simulate sensory variability. This setup allowed the networks to process essential quantitative information required for decision making without visual processing.

Offer Value Calculation. The value of each offer was calculated as:

$$\text{Offer Value} = \rho \times \text{Quantity} \times \text{Probability}, \tag{3}$$

where ρ is the intrinsic value of the good, Quantity is sampled uniformly from 0 to $10/\rho$ (ensuring comparable value ranges across goods and non-trivial decisions), and Probability is sampled uniformly from 0 to 1 (set to 1 in tasks without probability). Offers were selected to have values greater than one.

Trial Outcomes and Rewards. In all tasks, the networks were required to maintain fixation during the fixation, rule cue, and offer presentation periods. Breaking fixation prematurely resulted in trial abortion and a negative reward of -1. During the response period, the networks could select an action corresponding to one of the offered goods or continue to maintain fixation. Selecting an action ended the trial, and the networks received a reward based on the intrinsic value and quantity of the chosen offer. In tasks involving probabilities, the reward was delivered probabilistically according to the offered probability; otherwise, the reward was deterministic.

Task Cues and Inputs. Each task was indicated to the networks via a unique rule cue presented during the rule cue period. The rule cues were distinct scalar inputs, allowing the networks to identify the current task and adjust their computations accordingly. This required the networks to develop task-dependent processing strategies akin to cognitive flexibility observed in animals performing multiple tasks [65].

Network Initialization. The hidden states of the recurrent networks were set to zero for each unit at the beginning of each trial for simplicity, ensuring that the network dynamics had no memory trace of the previous trials.

Below, we describe each task in detail, highlighting their specific features.

Standard Task

In the *standard task*, the networks were presented with two offers simultaneously, each consisting of a single good (either good C or good E) varying in quantity. Probabilities were not involved in this task (set to 1). The networks received scalar inputs for goods C and E quantities through dedicated input channels. The value of each offer was computed as:

$$\text{Offer Value} = \rho \times \text{Quantity}. \tag{4}$$

The networks had to select the offer with the higher value by comparing these values based on the inputs.

Risky Task

The *risky task* introduced probabilistic outcomes. Each offer, consisting of either good C or good E, varied in both quantity and probability. The networks received scalar inputs for each good's quantity and probability through dedicated input channels. The expected value of each offer was computed as:

$$\text{Offer Value} = \rho \times \text{Quantity} \times \text{Probability}. \tag{5}$$

The networks had to integrate the inputs to assess the expected values and choose the higher offer.

Bundles Task

In the *bundles task*, the networks chose between two bundles, each containing two different goods presented simultaneously. One bundle consisted of goods B and C and the other of goods D and E. Each good varied in quantity and probability. The networks received scalar inputs for the quantities and probabilities of each good through dedicated input channels. The total value of each bundle was calculated by summing the values of the individual goods:

$$\text{Bundle Value} = \rho_1 \times \text{Quantity}_1 \times \text{Probability}_1 + \rho_2 \times \text{Quantity}_2 \times \text{Probability}_2. \tag{6}$$

The networks had to process multiple input channels corresponding to different goods and perform additive computations to determine the overall bundle values.

Ternary Task

The *ternary task* involved choosing among three different goods (goods A, C, and E), each varying in both quantity and probability. The networks received separate scalar inputs for the quantity and probability of each good through dedicated input channels. The expected value of each offer was computed as:

$$\text{Offer Value} = \rho \times \text{Quantity} \times \text{Probability}. \tag{7}$$

The networks had to evaluate and compare the expected values based on the inputs and select the offer with the highest expected value.

Sequential Task

Two offers were presented sequentially in the *sequential task*, with variable delays between presentations. Each offer consisted of a single good (either good C or good E) varying in quantity and probability. The sequence of presentation was randomized. During each offer presentation, the networks received scalar inputs for the quantity

and probability of the presented good through dedicated input channels. The networks needed to maintain the value of the first offer during the delay period. Upon presentation of the second offer, the networks had to compare it with the stored representation to decide which one had the higher value:

$$\text{Offer Value} = \rho \times \text{Quantity} \times \text{Probability}. \tag{8}$$

This task tested the networks' working memory capabilities, requiring retention and comparison of information over delays based on the scalar inputs.

By designing and training the networks on these detailed and varied tasks, we aimed to simulate the complexity of decision processes observed in biological systems. The careful mapping of visual representations to scalar inputs allowed the networks to process essential quantitative information required for decision making, focusing on quantities, probabilities, and goods identity through dedicated input channels. This approach facilitated the investigation of how neural networks can develop mechanisms to perform value-based decisions involving multiple variables and to explore the potential neural correlates of such processes.

Training procedure

We trained the networks using Proximal Policy Optimization (PPO) [21, 66], a state-of-the-art deep reinforcement learning algorithm. PPO belongs to the family of policy gradient algorithms, which optimize the policy directly through gradient ascent, focusing on maximizing the expected cumulative reward. Other algorithms in this family include REINFORCE and Advantage Actor-Critic (A2C), which have been successfully applied to study neuroscience problems before [15, 49].

We defined the loss function $\mathcal{L}(\theta)$ to be maximized on every training batch of trials as a weighted sum of the PPO policy loss $\mathcal{L}_t^{\text{PPO}}(\theta)$, the value function loss $\mathcal{L}_t^{\text{VF}}(\theta)$, and an entropy regularization term $S[\pi_\theta](s_t)$:

$$\mathcal{L}(\theta) = \mathbb{E}\left[\mathcal{L}_t^{\text{PPO}}(\theta) - c_1\mathcal{L}_t^{\text{VF}}(\theta) + c_2 S[\pi_\theta](s_t)\right], \tag{9}$$

where $c_1 = 0.5$ and $c_2 = 0.01$ are hyperparameters determining the weights of the value function loss and the entropy regularization term, respectively. $\mathbb{E}$ represents the mean over a batch of training trials composed by unrolling 20 environments, simulated in parallel with the same agent, for $T = 128$ steps. This batch was then split into 4 mini-batches used for optimization. θ refers to the collection of all trainable parameters, and π_θ is the policy used to sample the actions of the network given an input.

The policy loss $\mathcal{L}_t^{\text{PPO}}$ is defined as:

$$\mathcal{L}_t^{\text{PPO}}(\theta) = -\min\left(\rho_t(\theta)A_t,\, \text{clip}\left(\rho_t(\theta), 1-\epsilon, 1+\epsilon\right)A_t\right), \tag{10}$$

where $\rho_t(\theta) = \frac{\pi_\theta(a_t|s_t)}{\pi_\theta^{\text{old}}(a_t|s_t)}$ is the probability ratio between the current and old policies, $\epsilon = 0.1$ is the clipping parameter, and A_t is the advantage function estimating the relative value of action a_t at state s_t.

The advantage function A_t (analogous to the reward prediction error in neuroscience) is defined as:

$$A_t = -V_\theta(s_t) + r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots + \gamma^{T-t} r_{T-1} + \gamma^{T-t} V_\theta(s_T), \tag{11}$$

where r_t is the actual reward at time step t, $\gamma = 0.99$ is the temporal discount factor, $V_\theta(s_t)$ is the value function computed at state s_t, and T is the time horizon. The state $s_t = \mathbf{u}_t$ is the input to the network at time t. This advantage function represents the difference between the cumulative future rewards (including the bootstrap from the value function at time T) and the value estimate at time t.

The value function loss was:

$$\mathcal{L}_t^{\mathrm{VF}}(\theta) = \frac{1}{2}\left(V_\theta(s_t) - R_t\right)^2, \tag{12}$$

where R_t is the bootstrapped return computed as:

$$R_t = r_t + \gamma r_{t+1} + \gamma^2 r_{t+2} + \ldots + \gamma^{T-t} r_{T-1} + \gamma^{T-t} V_\theta(s_T). \tag{13}$$

Alternatively, we can express the bootstrapped return as:

$$R_t = A_t + V_\theta(s_t). \tag{14}$$

The entropy bonus was:

$$S[\pi_\theta](s_t) = -\sum_a \pi_\theta(a|s_t) \log \pi_\theta(a|s_t), \tag{15}$$

encouraging exploration by promoting a wider distribution of actions.

In this context, the advantage function A_t serves a role analogous to the reward prediction error (RPE) in neuroscience [67, 68]. The RPE represents the discrepancy between the expected reward and the actual obtained reward, driving learning and adaptation in biological neural systems. Similarly, the advantage function measures the difference between the estimated value of the current state $V_\theta(s_t)$ and the actual cumulative future rewards, guiding the adjustment of the policy to maximize expected returns.

All parameters were updated via gradient ascent and backpropagation through time using the Adam optimizer with a learning rate of 2.5×10^{-4} [69]. We clipped the gradient norm to be less than or equal to 1 to prevent exploding gradients.

Networks were trained under both multitasking and curriculum learning protocols [28, 30]. In multitasking, networks were trained concurrently on all tasks, facilitating generalization across tasks. In curriculum learning, networks were first trained on simplified versions of tasks and gradually introduced to more complex variants, aiding convergence and performance. For example, when training on the sequential task, networks were initially trained without delays, and delays were gradually introduced once the networks could solve the simplified task.

Training continued until the networks met stringent performance criteria evaluated on test trials: achieving at least 99% of decision trials (trials completed without fixation breaks) (Fig. S1**a**) and at least 90% correct choices among those decision trials (Fig. S1**b**). The learning curves reveal distinct phases: an initial phase where the networks learn to maintain fixation and choose randomly between goods, followed by a phase where they learn to choose the good with the highest value. The average reward on the test set increased correspondingly (Fig. S1**c**), reflecting improved decision making performance.

All simulations were implemented using PyTorch [70] and custom Python scripts. Performance metrics and network dynamics were analyzed post-training.

Logistic Regression Analysis of Behavior

To quantitatively characterize the networks' decision making and extract behavioral parameters, we performed logistic regression analyses on the choice data from each task, following methodologies applied in studies with non-human primates [7].

Data Collection

For each network, we collected approximately 25,000 trials (about 5,000 trials per task) where the network made choices among different offers. Each offer was defined by its quantity (q_X), probability (p_X) when applicable, and intrinsic value (ρ_X). The intrinsic values were set during training, with $\rho_A > \rho_B > \rho_C > \rho_D > \rho_E$.

Models for Binary Choices

In tasks involving choices between two options—the *standard*, *risky*, *bundles*, and *sequential* tasks—we modeled the probability of choosing one option over another using logistic regression. The models estimate the relative values ρ_X of goods relative to a reference good, which we designate as good E. By convention, we set $\rho_E = 1$, and ρ_X represents the relative value of good X compared to good E.

Standard Task

In the *standard task*, choices are based on quantities and intrinsic values of goods C and E, without probabilities. The probability of choosing good C over good E is modeled as:

$$P_{\text{choose C}} = \frac{1}{1 + \exp\left(-\eta\left(\ln\left(\frac{\rho_C q_C}{q_E}\right)\right)\right)}, \tag{16}$$

where:

- $P_{\text{choose C}}$ is the probability of choosing good C.
- ρ_C is the relative value of good C relative to good E ($\rho_E = 1$).
- q_C and q_E are the quantities offered.
- η is the choice consistency parameter, reflecting sensitivity to value differences.

Risky Task

In the *risky task*, choices involve quantities, probabilities, and intrinsic values. We included the risk attitude parameter γ to capture potential non-linear weighting of probabilities:

$$P_{\text{choose C}} = \frac{1}{1 + \exp\left(-\eta\left(\ln\left(\frac{\rho_C q_C p_C^\gamma}{q_E p_E^\gamma}\right)\right)\right)}, \tag{17}$$

where:

- p_C and p_E are the probabilities associated with goods C and E.
- γ quantifies the network's risk attitude ($\gamma = 1$ denotes risk neutrality).
- Other parameters are as previously defined.

Bundles Task

In the *bundles task*, each option is a bundle consisting of two goods. We calculated the total value of each bundle by summing the values of the individual goods. The probability of choosing Bundle 1 over Bundle 2 is modeled as:

$$P_{\text{choose Bundle 1}} = \frac{1}{1 + \exp\left(-\eta\left(\ln\left(\frac{V_1}{V_2}\right)\right)\right)}, \tag{18}$$

where:

- $V_1 = \rho_B q_B p_B^\gamma + \rho_C q_C p_C^\gamma$ is the total value of Bundle 1 (goods B and C).
- $V_2 = \rho_D q_D p_D^\gamma + q_E p_E^\gamma$ is the total value of Bundle 2 (goods D and E, with $\rho_E = 1$).
- γ is the risk attitude parameter.
- ρ_X, q_X, and p_X are the intrinsic values, quantities, and probabilities of the goods in the bundles.
- η is the choice consistency parameter.

Sequential Task

In the *sequential task*, choices involve quantities, probabilities, intrinsic values, and potential order bias. The probability of choosing good C over good E is modeled as:

$$P_{\text{choose C}} = \frac{1}{1 + \exp\left(-\eta\left(\ln\left(\frac{\rho_C q_C p_C^\gamma}{q_E p_E^\gamma}\right) + \epsilon' \cdot \text{Order}\right)\right)}, \tag{19}$$

where:

- ϵ' is the order bias parameter, capturing preference for the first or second offer.
- Order $= +1$ if good C was presented second, -1 if good C was presented first.
- Other parameters are as previously defined.

The risk attitude parameter γ is included due to the involvement of probabilities.

Model for Multinomial Choices

In the *ternary task*, involving choices among three goods (A, C, and E), we used multinomial logistic regression to model the probability of choosing each good based on its offer value:

$$P_{\text{choose } X} = \frac{(\rho_X q_X p_X^\gamma)^\eta}{\sum_Y (\rho_Y q_Y p_Y^\gamma)^\eta}, \tag{20}$$

where:

- X and Y index the goods offered (A, C, E).
- ρ_X is the relative value of good X relative to good E ($\rho_E = 1$).
- q_X and p_X are the quantity and probability of good X.
- γ is the risk attitude parameter.
- η is the choice consistency parameter.

Parameter Estimation

We estimated the parameters (ρ_X, γ, η, ϵ') using maximum likelihood estimation for each task and network individually. The relative values (ρ_X) are relative measures, with $\rho_E = 1$ serving as the reference point, allowing for comparison across tasks and networks. The choice consistency parameter (η) indicates the steepness of the psychometric function; higher values correspond to more consistent choices based on offer value differences.

Single-neuron analyses

Data Collection and Trial Structure

For the neural analyses, we collected test trials with fixed durations, set to the maximum durations for each epoch used during training. Specifically, we fixed the durations for fixation, rule cue, offer presentation, and response periods to their maximum values used during training. This standardization facilitated the alignment of neural activity across trials and simplified the temporal analysis of neuronal selectivity.

In the *sequential task*, when analyzing neural activity aligned to stimulus onset, we refer specifically to the onset of the *second* stimulus. At this point in the trial, the network has information about the first offer maintained in working memory and receives the second offer, enabling it to compare both offers to make a decision. By focusing on the period following the second stimulus onset, we capture the neural processes involved in integrating sequential information to guide choice behavior.

Single-neuron selectivity analysis

To analyze neuronal selectivity for decision variables, we performed linear regression of each neuron's firing rate against each behavioral variable independently at each time point during the trial. The considered variables included:

- Offer value of good C (OVC)
- Offer value of good E (OVE)
- Chosen value (CV)
- Choice (CH), coded as +1 for choosing good C and -1 for choosing good E
- Value sum (sum of the offer values)
- Value difference (difference between the offer values)

For each neuron and time point, we fitted the following model separately for each variable:

$$r(t) = \beta_0 + \beta X + \epsilon, \tag{21}$$

where $r(t)$ is the firing rate of the neuron at time t, X is the behavioral variable, β is the regression coefficient, and ϵ is the error term.

We evaluated the statistical significance of each regression by examining the p-value associated with the regression coefficient β. If the p-value was less than 0.05, we considered the regression significant; otherwise, the R^2 value was set to zero. At each time point, a neuron was assigned to the variable with the highest R^2 among those with significant regressions. If none of the regressions were significant ($p \geq 0.05$), the neuron was considered non-selective at that time point.

Temporal Stability Index (TSI)

The TSI for each neuron was calculated as:

$$\text{TSI} = \frac{\text{Number of times the neuron encodes its primary variable}}{\text{Total number of times the neuron is selective}}, \tag{22}$$

where the primary variable is the one most frequently associated with the neuron over time. A TSI of 1 indicates that the neuron consistently encodes the same variable whenever it is selective, while lower values suggest that the neuron's selectivity changes over time.

Note that a TSI value below 0.5 indicates that a neuron frequently shifts its primary encoded variable during a trial. This suggests that such neurons exhibit dynamic selectivity, rather than consistently encoding a single decision variable.

Categorical encoding analysis

To assess whether neurons encode variables categorically or conjunctively, we focused on neurons that were selective for at least one of the variables at a given time (i.e., had $R^2 \geq 0.3$ for at least one variable). For each neuron and time point, we computed the difference in R^2 values between pairs of decision variables. Specifically, we calculated:

$$\Delta R^2 = R^2_{\text{Var1}} - R^2_{\text{Var2}}, \tag{23}$$

where R^2_{Var1} and R^2_{Var2} are the coefficients of determination from the independent regressions against variables Var1 and Var2, respectively. If a regression was not significant ($p \geq 0.05$), we set R^2 to zero. Histograms of these differences were plotted to examine the distribution across the neuronal population. Bimodal distributions suggest categorical encoding, where neurons preferentially encode one variable over another, whereas unimodal distributions indicate conjunctive encoding. We opted for pairwise comparisons of R^2 values between candidate variables as this method provides a simple and interpretable metric for determining which variable dominates the neuronal response.

Neural Population Analyses

Population Dynamics Analysis

To investigate the population-level encoding of decision variables, we performed principal component analysis (PCA) on the neural activity of the recurrent neurons during the stimulus presentation phase. PCA reduces the high-dimensional neural activity to a set of orthogonal components that capture the maximum variance in the data.

For each task and network, we collected neural activity from the last 200 ms of the stimulus presentation phase across all trials. We separately analyzed the excitatory and inhibitory populations to examine potential differences in their contributions to the population dynamics.

Dimensionality Estimation

We estimated the dimensionality of the neural activity using the participation ratio [35], defined as:

$$D_{\mathrm{PR}} = \left(\sum_i \lambda_i\right)^2 \Big/ \sum_i \lambda_i^2, \tag{24}$$

where λ_i are the eigenvalues (variance explained) of the covariance matrix of the neural activity. The participation ratio provides a measure of the effective number of dimensions contributing to the variance in the data.

Principal Component Regression

We projected the neural activity onto the first three principal components and performed linear regression of these projections against various decision variables, including offer values, chosen value, choice, value sum, and value difference. The coefficients of determination (R^2) from these regressions indicate how much variance in each principal component is explained by the decision variables.

Lesion Analysis

To assess the role of recurrent connectivity in shaping population dynamics, we performed lesion experiments by removing all recurrent connections from the trained networks. Specifically, we set the recurrent weight matrix W^{rec} to zero, effectively eliminating all recurrent influences while preserving the feedforward inputs and output weights.

We then re-evaluated the networks' neural activity during the stimulus presentation phase using the same input protocols as in the intact networks. Since the networks could no longer perform the tasks without recurrent dynamics, we focused on the encoding of offer values and other input-related variables.

Firing Rate Distribution

We compared the distribution of firing rates between the intact and lesioned networks to assess the impact of recurrent connectivity on neuronal activity levels. Mean firing rates and standard deviations were computed for excitatory and inhibitory neurons across all tasks (Supplementary Fig. S4).

Population Dynamics in Lesioned Networks

We repeated the PCA and dimensionality analyses on the lesioned networks to examine how the absence of recurrent connections affected the population encoding of decision variables. Linear regression of the principal components against offer values and other variables was performed to determine the nature of the encoding in the feedforward regime (Supplementary Fig. S5).

Visualization of Population Dynamics

To visualize the temporal evolution of the population activity, we created animations showing the trajectories of neural activity projected onto the first two principal components throughout the trial. Trials were colored according to different decision variables to illustrate how neural trajectories corresponding to different choices or offer values diverged over time.

Supplementary Video 1 shows the population dynamics in the intact network during the risky task, highlighting the separation of trajectories based on chosen value and choice.

Supplementary Video 2 shows the population dynamics in the lesioned network during the risky task, illustrating the lack of separation based on choice-related variables and the clustering of trajectories according to offer values.

Input Weights Analyses

Analysis of Input Weights

We examined the input weight matrices W^{in} connecting the input units for quantities and probabilities to the recurrent neurons. For each network, we extracted the input weight vectors corresponding to each feature (quantity and probability) of each good. We then computed the Pearson correlation coefficients between all pairs of these input weight vectors to assess the similarity in how recurrent neurons receive inputs from different features and goods. The resulting correlation matrix allowed us to identify patterns indicating the specialized processing of goods by subpopulations of neurons.

Multiplicative vs. Additive Computation

We analyzed networks with lesioned recurrent connections to focus on feedforward computations to determine whether the networks are computing the offer values via multiplication or addition of quantities and probabilities. We collected neural activity from the last 200 ms of the stimulus presentation phase in the risky task across approximately 5,000 trials. We projected the population activity of excitatory and inhibitory neurons onto the first two principal components using principal component analysis (PCA). We then performed linear regression of these projections against both the product (multiplication) and the sum of quantities and probabilities for each good. The coefficients of determination (R^2) were compared to assess which model better explained the neural activity.

Toy Feedforward Network Model

We constructed a simplified feedforward network model to illustrate how multiplication can be approximated using linear weights and nonlinear activation functions (ReLU). In our first step, we optimized the scaling parameters α (for the input weights) and β (for the biases) in a single-product scenario—i.e., for one good—by generating a test set of input pairs (quantity and probability) and evaluating the ability of the network to approximate the product (multiplicative integration) versus the sum. The model consisted of input units for the quantity and probability of a single good, connected to hidden units with weights that scaled linearly with neuron indices. Bias terms were included to adjust the activation thresholds. The hidden units' activities were passed through ReLU functions, and principal component analysis (PCA) was performed on the hidden layer to analyze the population coding.

After optimizing α and β for this single-product case (see Supplementary Fig.S6), we generalized the model to the case of two goods. Initially, the model featured two independent populations of hidden units, each processing inputs from one of the goods.

Subsequently, we introduced a third population of mixed-selectivity neurons that integrated inputs from both goods. This extension allowed the model to approximate the multiplication operation for two goods, while also accounting for the observed rotation of the offer value axes in the full recurrent network (see Fig.5**c** and **d**).

Correlation of relative values and Input Weights

We trained 50 networks exclusively on the risky task, varying the intrinsic value ρ_{high} of the higher-value good across networks ($\rho_{\text{high}} = 1, 2, 3, 4, 5$). After training, we performed logistic regression to infer the relative values ρ_{high} for each network. We then calculated the average non-zero input weights connecting the quantity input of the high-value good to the recurrent neurons.

A linear regression was performed between the inferred relative values and the average non-zero input weights connecting the quantity input of the high-value good to the recurrent neurons to assess their relationship. A strong positive correlation would indicate that the relative values are encoded in the input weights, supporting the hypothesis that value computation occurs at the input level.

Recurrent Circuit Analyses

Analysis of Output Dynamics

We analyzed the activity of the network's output units during the stimulus presentation phase across all tasks and networks. The outputs included units corresponding to choices of goods A, C, E, fixation (FIX), and the expected return (value function). For each task, we averaged the output activity across trials where different choices were made, focusing on the dynamics leading up to the response phase. Error bars represent the standard error of the mean across networks.

Reaction Time Analysis

Reaction times (RTs) were measured from the onset of the response phase (when the fixation cue turned off) to the time the network executed a choice action. We collected RTs across all trials and tasks and analyzed their distributions (Supplementary Fig. S7**a**). Linear regression was performed between RTs and the absolute value difference between offers to assess the relationship between decision difficulty and response latency. Low coefficients of determination (R^2) indicated that RTs were not significantly influenced by the value difference.

Recurrent Connectivity Analysis

We examined the recurrent weight matrices W^{rec} of the trained networks. Singular value decomposition (SVD) was used to assess the rank and identify dominant connectivity patterns (Supplementary Fig. S7**b**). The number of singular values needed to explain 80% of the variance provided a measure of the low-rank structure of the connectivity.

Reduced Connectivity Matrix

Neurons were categorized based on their selectivity for decision variables (offer value, chosen value, choice) determined from the single-neuron selectivity analysis and their excitatory or inhibitory type. We constructed a reduced connectivity matrix by averaging the recurrent weights between these neuron groups across all networks. The matrix entries represent the mean synaptic strength from neurons in one category to another, highlighting key connectivity motifs underlying the network's dynamics (Fig. 6**c**). The analysis was performed separately for each task.

Circuit Diagram Visualization

Using the reduced connectivity matrix, we created simplified circuit diagrams to visualize the interactions between neuron populations. We focused on the choice and chosen value populations to illustrate the excitatory and inhibitory connections that facilitate winner-take-all dynamics and value comparison (Fig. 6**d**). The diagrams help conceptualize how the network's architecture supports the decision process.

Multitask and Curriculum Analyses

Rule Cue Input Weight Analysis

We analyzed the input weight matrices W^{in} associated with the rule cues for each task. For each network, we extracted the input weight vectors corresponding to the rule cue inputs and computed the Pearson correlation coefficients between all pairs of these vectors across tasks. Averaging these correlations over all networks provided insights into the similarity of rule cue representations and the extent of shared input structures among tasks.

Subspace Analysis of Population Activity

We performed principal component analysis (PCA) on the population activity during the rule cue period to examine the neural representations underlying each task. The participation ratio was calculated to estimate the dimensionality of the neural activity for each task. We then computed the angles between the subspaces spanned by the principal components of different tasks to assess the overlap in neural representations. Smaller angles indicate more overlapping subspaces, suggesting shared neural dynamics.

Task Variance and Neuronal Clustering

We calculated the variance of each neuron's firing rate during the stimulus presentation phase across all trial types. The variance for each task was normalized by the maximum variance observed for that neuron across tasks. Neurons were then clustered based on their normalized task variances using k-means clustering, with the optimal number of clusters determined by the silhouette score. This allowed us to identify groups of neurons contributing to shared or task-specific computations.

Fractional Task Variance Analysis

For each pair of tasks, we computed the difference in normalized firing rate variances divided by the sum for each neuron. We plotted histograms of these differences to visualize the relationships between tasks. Histograms with peaks at zero and one of the extremes indicate inclusive relationships, where one task's neural representation is a subset of another's. Histograms with multiple peaks suggest disjoint relationships, reflecting task-specific neuronal populations.

Visualization of Task Representations

We computed the mean firing rates of all neurons during the stimulus presentation phase for each task and projected these high-dimensional vectors onto the first three principal components. Aligning the principal components across networks, we visualized the clustering of task representations, highlighting the compositional structure of the network's neural activity.

Curriculum Learning Protocols

We implemented curriculum learning protocols in which networks were trained sequentially on tasks, starting with simpler tasks and progressing to more complex ones. To assess the benefits of prior learning and schema formation, we compared networks trained from scratch to those using curriculum learning.

Generalization Analyses

Training with Constrained Offer Sets

We trained ten networks exclusively on the risky task using a constrained set of offers. In the first training condition, the probability p was fixed at 0.5 (half of its maximum value), while the quantity q varied uniformly across its full range for both goods C and E. In the second condition, the quantity was fixed at half of its maximum value, and the probability varied uniformly between 0 and 1. This design ensured that during training, the networks encountered offers where only one variable (either q or p) changed, limiting their experience to a subset of possible offer combinations.

The networks were trained using the same reinforcement learning protocols described above. Training continued until the networks achieved high performance in selecting the higher-valued offer within the constrained offer set.

Testing Generalization with Unconstrained Offers

After training, we assessed the networks' ability to generalize by testing them on an unconstrained set of offers where both quantity and probability varied independently across their full ranges. This test set included offer combinations that the networks had not encountered during training, requiring them to compute offer values involving novel quantity-probability pairs.

We evaluated the networks' performance by measuring the percentage of correct choices—selecting the offer with the higher expected value—across a large number of test trials. We also analyzed choice patterns to assess whether the networks exhibited similar choice behavior in the test set compared to the training set.

Behavioral Analysis

We performed logistic regression analyses on the choice data from both the constrained training set and the unconstrained test set for each network. The logistic model was defined as:

$$P_{\text{choose C}} = \frac{1}{1 + \exp\left(-\eta\left(\ln\left(\frac{\rho_C q_C p_C^{\gamma}}{q_E p_E^{\gamma}}\right)\right)\right)}, \tag{25}$$

where ρ_C is the relative value of good C relative to good E (with $\rho_E = 1$), γ is the risk attitude parameter, η is the choice consistency parameter, and q_X, p_X are the quantities and probabilities of goods C and E, respectively.

By fitting the model to the choice data, we extracted the inferred relative values, risk attitudes, choice consistency, and accuracy for each network in both the training and test conditions. Comparisons of these parameters allowed us to assess the consistency of value computation and decision strategies between familiar and novel offer combinations.

Code availability

All training and analysis codes will be available on GitHub upon publication.

Data availability

Upon publication, we will provide data files in Python-readable formats for all trained models for further analysis on GitHub.

References

7. Padoa-Schioppa, C. Logistic analysis of choice data: A primer. *Neuron* **110,** 1615–1630 (2022).
9. Ballesta, S., Shi, W., Conen, K. E. & Padoa-Schioppa, C. Values encoded in orbitofrontal cortex are causally related to economic choices. *Nature* **588,** 450–453 (2020).
15. Song, H. F., Yang, G. R. & Wang, X. Reward-based training of recurrent neural networks for cognitive and value-based tasks. *Elife* **6,** e21492 (2017).
17. Wang, X.-J. Probabilistic decision making by slow reverberation in cortical circuits. *Neuron* **36,** 955–968 (2002).
19. Song, H. F., Yang, G. R. & Wang, X. Training excitatory-inhibitory recurrent neural networks for cognitive tasks: a simple and flexible framework. *PLoS computational biology* **12,** e1004792 (2016).
21. Schulman, J., Wolski, F., Dhariwal, P., Radford, A. & Klimov, O. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347* (2017).
28. Yang, G. R., Joglekar, M. R., Song, H., Newsome, W. T. & Wang, X. Task representations in neural networks trained to perform many cognitive tasks. *Nature neuroscience* **22,** 297–306 (2019).
30. Bengio, Y., Louradour, J., Collobert, R. & Weston, J. *Curriculum learning* in *Proceedings of the 26th annual international conference on machine learning* (2009), 41–48.
35. Gao, P. *et al.* A theory of multineuronal dimensionality, dynamics and measurement. *BioRxiv,* 214262 (2017).
49. Wang, J. X. *et al.* Prefrontal cortex as a meta-reinforcement learning system. *Nature neuroscience* **21,** 860–868 (2018).
62. Markram, H. *et al.* Interneurons of the neocortical inhibitory system. *Nature reviews neuroscience* **5,** 793–807 (2004).
63. Douglas, R. J. & Martin, K. A. Recurrent neuronal circuits in the neocortex. *Current biology* **17,** R496–R500 (2007).
64. Rajan, K., Abbott, L. & Sompolinsky, H. Stimulus-dependent suppression of chaos in recurrent neural networks. *Physical Review E—Statistical, Nonlinear, and Soft Matter Physics* **82,** 011903 (2010).
65. Stoet, G. & Snyder, L. H. Neural correlates of executive control functions in the monkey. *Trends in cognitive sciences* **13,** 228–234 (2009).
66. Huang, S. *et al.* CleanRL: High-quality Single-file Implementations of Deep Reinforcement Learning Algorithms. *Journal of Machine Learning Research* **23,** 1–18. http://jmlr.org/papers/v23/21-1342.html (2022).
67. Schultz, W., Dayan, P. & Montague, P. R. A neural substrate of prediction and reward. *Science* **275,** 1593–1599 (1997).
68. Glimcher, P. W. Understanding dopamine and reinforcement learning: the dopamine reward prediction error hypothesis. *Proceedings of the National Academy of Sciences* **108,** 15647–15654 (2011).
69. Kingma, D. & Ba, J. Adam: A method for stochastic optimization. *arXiv preprint arXiv:1412.6980* (2014).
70. Paszke, A. *et al.* Pytorch: An imperative style, high-performance deep learning library. *Advances in neural information processing systems* **32** (2019).

Supplementary Figures

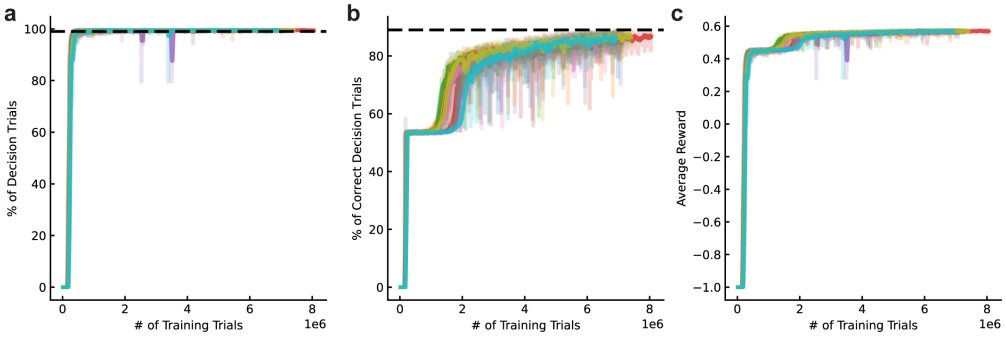

Fig. S1 Learning curves for networks trained across all tasks. a, Percentage of decision trials (*i.e.*, trials completed without fixation breaks) on the test set as a function of training trials. The black dashed line indicates the learning criterion of 99% decision trials. **b**, Percentage of correct choices among decision trials on the test set as a function of training trials. The black dashed line indicates the learning criterion of 90% correct decisions. **c**, Average reward on the test set as a function of training trials. In all panels, percentages and average rewards are computed on test trials. Colored lines represent different networks with random initializations; solid lines are smoothed versions. The learning curves show an initial phase where the networks learn to maintain fixation and choose randomly between goods, followed by a phase where they learn to select the highest-value offer. Training stops once both criteria are met.

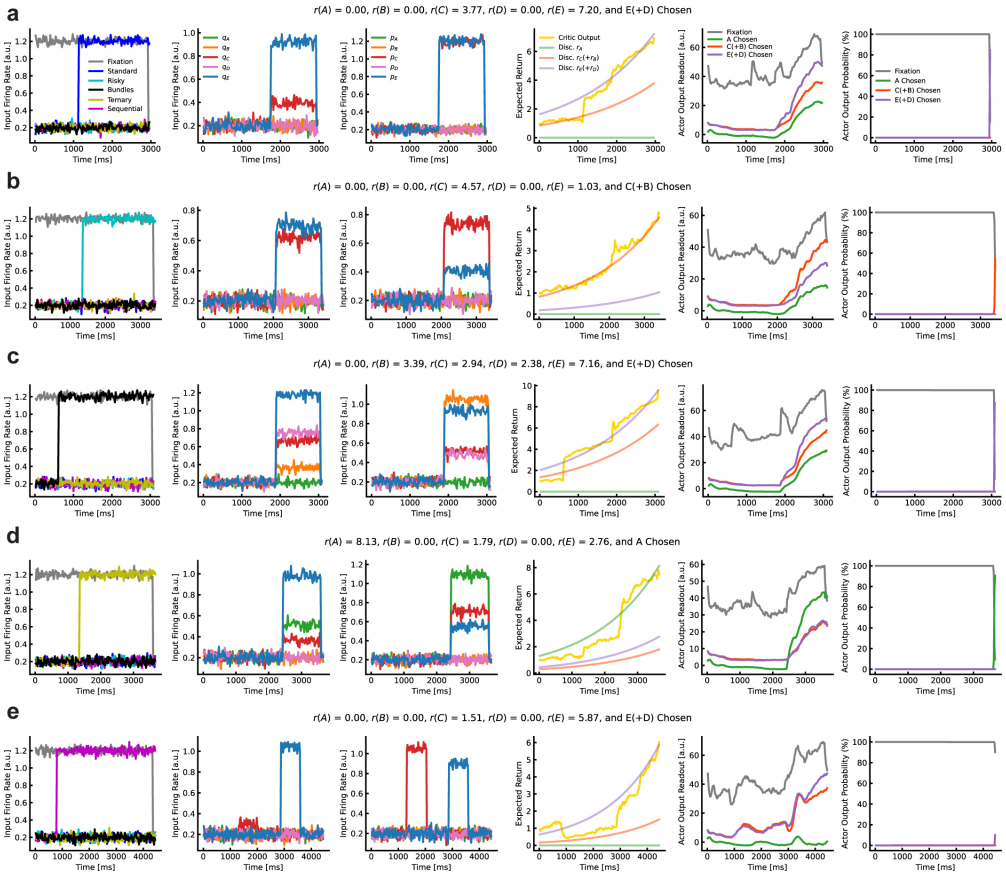

Fig. S2 Example trials from a trained network on the five economic choice tasks. Each panel shows the temporal dynamics of inputs and outputs for a single trial. The first column displays fixation and rule cue inputs over time. The second and third columns represent quantities and probabilities of the offered goods, respectively. The fourth column shows the value function output and temporally discounted potential rewards. The fifth and sixth columns illustrate policy outputs and softmax probabilities for different actions. **a**, Standard task. **b**, Risky task. **c**, Bundles task. **d**, Ternary task. **e**, Sequential task. In each trial, the network selects the highest-value offer. The value function predicts the expected return shortly after accessing the offers. Policy outputs are dominated by fixation during the fixation period, preventing premature selection of one of the offers, but the forthcoming choice can be inferred during the offer presentation phase. The trials have varying temporal durations, reflecting the variability in the task design.

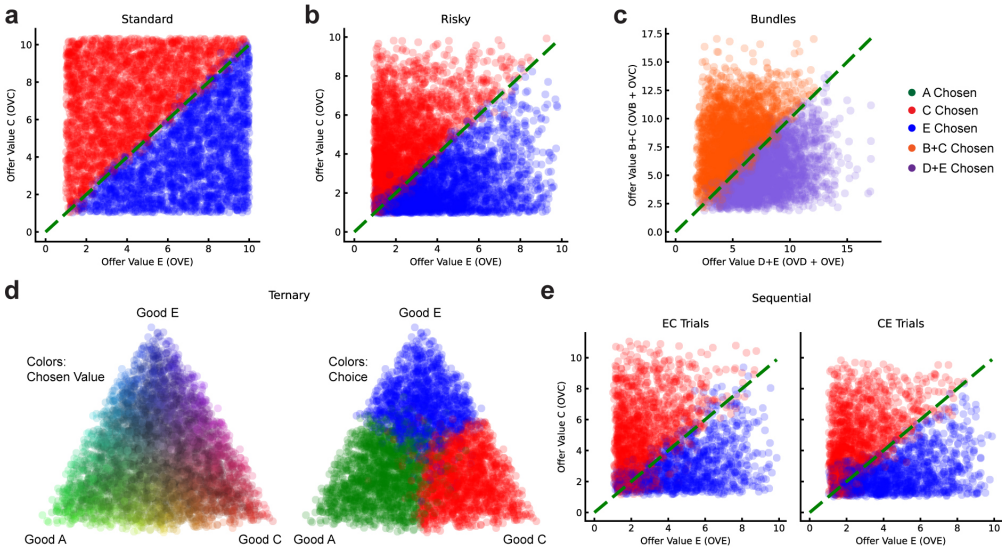

Fig. S3 Choice patterns of a network trained on all economic choice tasks. Each panel displays the network's choices in offer value space for a different task. Each point represents a trial, with axes corresponding to the offer values of the goods and colors indicating the network's choice. **a**, *Standard task*: The network consistently chooses the offer with the higher computed value based on quantity and intrinsic value. **b**, *Risky task*: As in Figure 2**a**, the network's choices are linearly separable in offer value space. **c**, *Bundles task*: The network computes the total value of each bundle by summing the values of the constituent goods and selects the bundle with the higher total value. **d**, *Ternary task*: Trials are plotted using a simplex (equilateral triangle) where each vertex represents the scenario where one of the three goods (A, C, or E) has the highest offer value while the other two have zero or lower values. The center of the triangle corresponds to trials where all three goods have equal offer values. Positions within the simplex reflect the relative offer values of the goods in each trial; points closer to a vertex indicate a higher offer value for that specific good. We present two simplex plots for the ternary task: **Left plot**: Points are colored based on the *offer values* of the goods using an RGB color scheme, where each color channel corresponds to one good (green for good A, red for good C, blue for good E). The intensity of each color channel reflects the magnitude of the offer value for that good in the trial. Thus, the color of each point visually represents the combination of offer values, with the dominant color indicating the good with the highest offer value. **Right plot**: The same trials are plotted, but points are colored according to the *network's choice* (green if the network chose good A, red for good C, blue for good E). This allows for a direct comparison between the offer values and the network's decisions. By comparing the two plots, it is evident that the network reliably chooses the good with the highest offer value across trials. Points that are colored similarly in both plots confirm that the network's choice aligns with the good that has the highest computed offer value in that trial. **e**, *Sequential task*: Separate plots for trials where good E is presented first or second. The network maintains information about the first offer and compares it with the second to choose the higher-value offer.

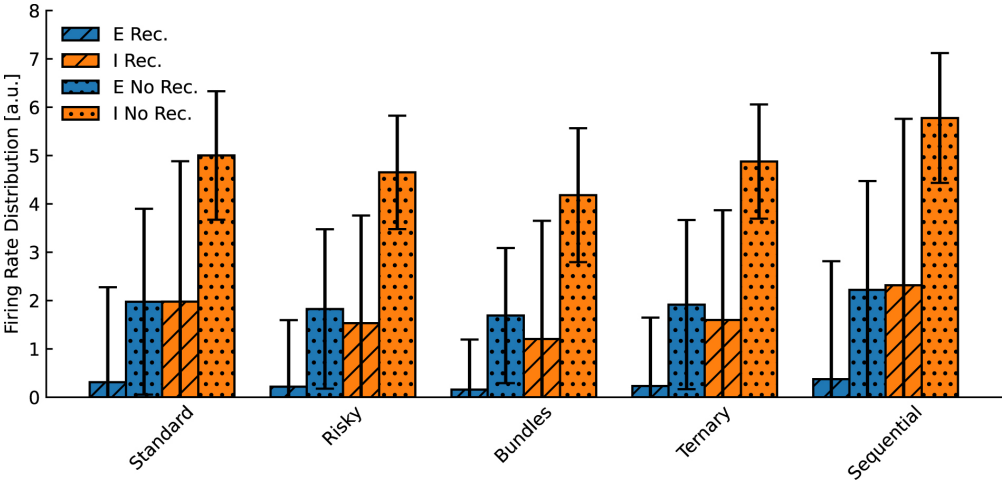

Fig. S4 Effect of lesioning recurrent connections on the distribution of firing rates in recurrent neurons across networks trained on all economic choice tasks. Bar plots show the mean firing rates of excitatory (blue) and inhibitory (orange) neurons during the stimulus presentation phase, comparing the intact networks (striped bars) to the lesioned networks without recurrent connections (dotted bars) across different tasks. Error bars represent the standard deviation. The removal of recurrent connections leads to an overall increase in firing rates due to the loss of inhibitory feedback and recurrent regulation of activity levels.

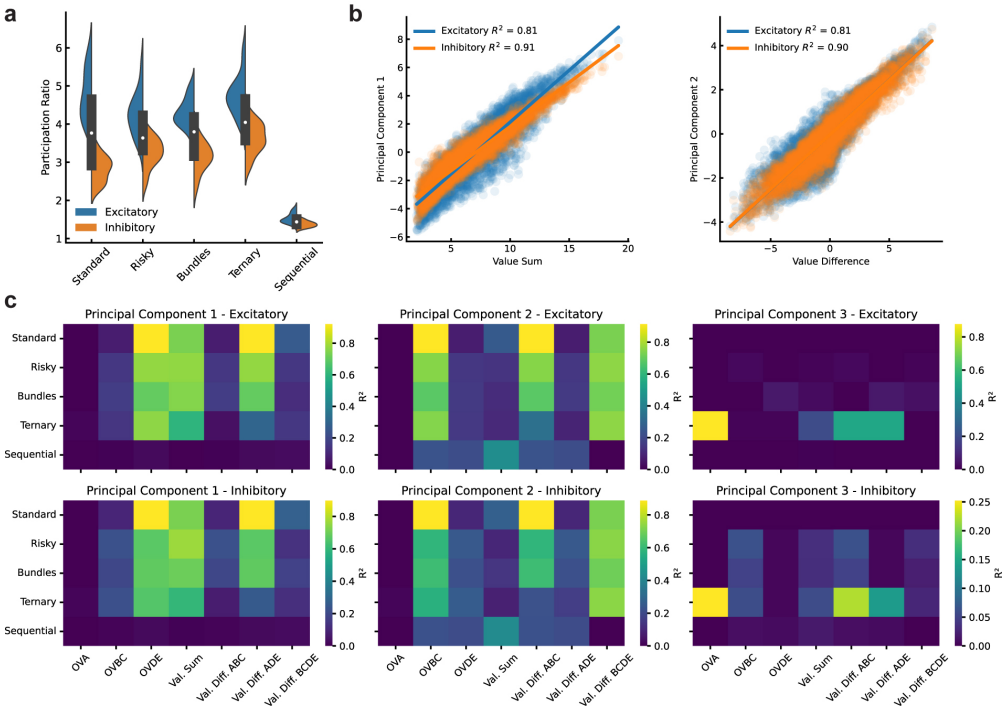

Fig. S5 Population analysis of networks trained on all economic choice tasks after lesioning all recurrent connections. a, Neural dimensionality in lesioned networks across different tasks, measured by the participation ratio. The dimensionality increases in most tasks compared to the intact networks, indicating that recurrent connectivity constrains neural activity into lower-dimensional manifolds. In the sequential task, dimensionality decreases due to the inability to maintain working memory without recurrence. **b**, PCA of neural activity during the risky task in a representative lesioned network. The first principal component (PC1) primarily encodes the value sum, while the second component (PC2) encodes the value difference. Each point represents a trial, colored according to the encoded variable. **c**, Summary of population analyses across all lesioned networks and tasks. Heatmaps display the average R^2 values from linear regression of the principal components onto offer values and related variables. The analyses indicate that without recurrent dynamics, the networks encode input-related variables but cannot represent chosen value or choice, underscoring the importance of recurrence for value comparison and decision making.

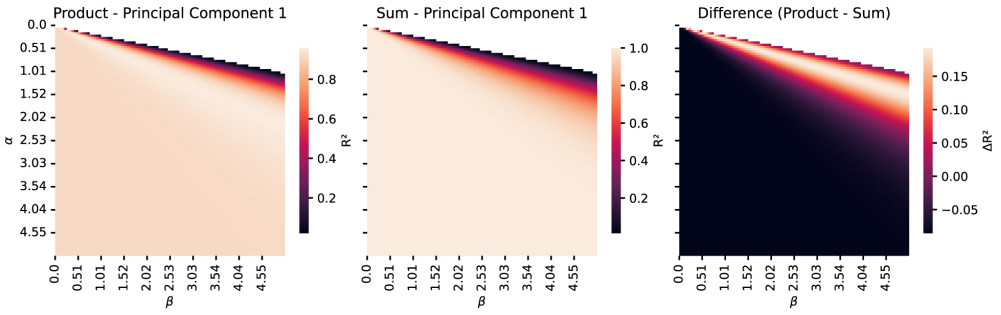

Fig. S6 Parameter search for the toy multiplication model. Heatmaps showing the coefficients of determination (R^2) from linear regressions of the first principal component of the hidden units against the product (left panel) and sum (middle panel) of inputs for different scaling parameters α (input weight scaling) and β (bias scaling). The right panel shows the difference between the R^2 values for the product and sum regressions, highlighting the parameter regions where the model better approximates multiplication over addition. Optimal parameters are those with higher R^2 for the product and a positive difference between the product and sum R^2.

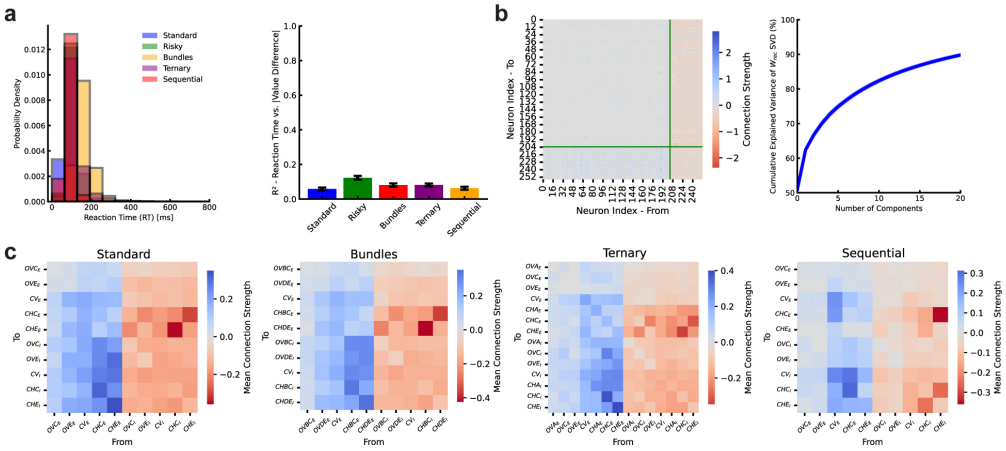

Fig. S7 Additional analyses of choice mechanisms and connectivity across economic choice tasks. a, Left: Distributions of reaction times (RTs) across all trained networks, separated by task. The RT distributions are similar across tasks, indicating consistent decision execution timing. Right: Linear regression of RTs against the absolute value difference between offers shows low coefficients of determination (R^2), suggesting no significant relationship between RT and decision difficulty. **b**, Left: Example of a full recurrent connectivity matrix from a trained network, illustrating its sparsity. Right: Singular value decomposition (SVD) analysis across all recurrent weight matrices, showing the number of singular values needed to explain 80% of the variance. The low number of significant singular values indicates a low-rank structure in the connectivity. **c**, Reduced connectivity matrices for other tasks, constructed similarly to Fig. 6**c**. Consistent winner-take-all motifs are observed across tasks, suggesting a common mechanism for value comparison implemented by the recurrent network.