Neural structure mapping in human probabilistic reward learning

F. Luyckx[1*], H. Nili[1,2], B. Spitzer[1,3†], C. Summerfield[1*†]

[1]Department of Experimental Psychology, University of Oxford, Oxford, OX2 6GG.

[2]Wellcome Centre for Integrative Neuroimaging, University of Oxford, Oxford OX3 9DU.

[3]Center for Adaptive Rationality, Max Planck Institute for Human Development, Berlin, Germany.

*Correspondence to: fabrice.luyckx@psy.ox.ac.uk, christopher.summerfield@psy.ox.ac.uk

†These authors contributed equally to this work.

Abstract

Humans can learn abstract concepts that describe invariances over relational patterns in data. One such concept, known as magnitude, allows stimuli to be compactly represented by a single dimension (i.e. on a mental line), for example according to their cardinality, size or value. Here, we measured representations of magnitude in humans by recording neural signals whilst they viewed symbolic numbers. During a subsequent reward-guided learning task, the neural patterns elicited by novel complex visual images reflected their pay-out probability in a way that suggested they were encoded onto the same mental number line. Our findings suggest that in humans, learning about values is accompanied by structural alignment of value representations with neural codes for the concept of magnitude.

The ability to learn rapidly from limited data is a key ingredient of human intelligence. For example, on moving to a new city, you will rapidly discover which restaurants offer good food and which neighbors provide enjoyable company. Current models of learning propose that appetitive actions towards novel stimuli are learned *tabula rasa* via reinforcement (*1*), and these models explain the amplitude of neural signals in diverse brain regions during reward-guided choices in humans and other animals (*2–4*). However, reinforcement learning models learn only gradually, and even when coupled with powerful function approximation methods, exhibit limited generalization beyond their training domain (*5*), leading to the suggestion they are ill-equipped to fully describe human learning (*6*). By contrast, cognitive scientists have ascribed human intelligence to formation of abstract knowledge representations (or "concepts") that delimit the structural forms that new data is likely to take (*7–9*). Indeed, real-world data can often be described by simple relational structures, such as a tree, a grid or a ring (*7*), and humans may infer relational structure through probabilistic computation (*10*) and understand new domains by their alignment with existing relational structures (*11*). However, these models are often criticized for failing to specify how concepts might be plausibly encoded or computed in neural circuits (*12*). A pressing concern, thus, is to provide a mechanistic account of how relational knowledge is encoded and generalized in the human brain (*13*).

The current project was inspired by recent observations that the representational geometry of human neural signals evoked by symbolic numbers respects their relative cardinality (*14*, *15*). In scalp M/EEG signals, neural patterns evoked by Arabic digits vary continuously with numerical distance, such that multivariate signals for "3" are more similar to those for "4" than "5". Number is a symbolic system that expresses magnitude in abstract form (*16–18*) and so we reasoned that

continuously-varying neural signals evoked by numbers might be indexing a conceptual basis set that supports one-dimensional encoding of novel stimuli. In the domain of reward-guided learning, a compact description of the stimulus space projects data into a single dimension that runs from "bad" to "good". Here, thus, we asked humans to learn the reward probabilities associated with novel, high-dimensional visual images, and measured whether the stimuli come to elicit neural patterns that map onto one-dimensional neural codes for numerical magnitude.

Whilst undergoing scalp EEG recordings, human participants (n = 46) completed two tasks: a numerical decision task and a probabilistic reward-guided learning task. In the numerical task participants viewed rapid streams of ten Arabic digits (1 to 6) and reported whether numbers in orange or blue font had the higher (n = 22) or lower (n = 24) average (Fig. 1A). Using representational similarity analysis (RSA) (*19*) we replicated the previous finding (*14*, *15*) that patterns of neural activity across the scalp from ~100ms onwards were increasingly dissimilar for digits with more divergent magnitude, i.e. codes for "3" and "5" were more dissimilar than those for "3" and "4" (Fig. 1A, green line). This occurred irrespective of task framing (report higher vs. lower average) and category (orange vs. blue numbers), suggesting that neural signals encoded an abstract representation of magnitude and not solely a decision-related quantity such as choice certainty (*14*). The reward-learning task was based on the multi-armed bandit paradigm that has been used ubiquitously to study value-guided decision-making (*20*). Participants learned the reward probabilities associated with six unique novel images (colored donkeys; Fig. 1A), which paid out a fixed reward with a stationary probability (range 0.05-0.95). These probability values were never signaled to the participant but instead acquired by trial and error

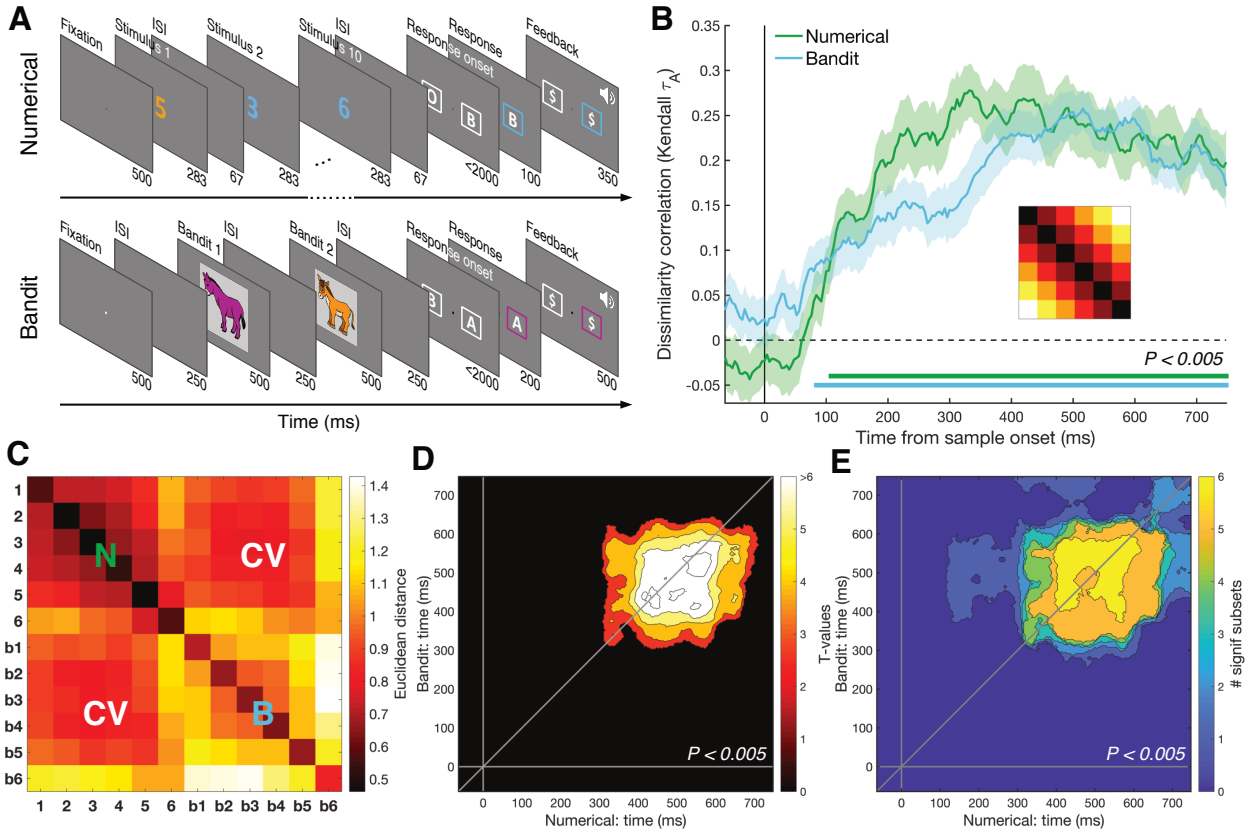

Fig. 1. *Task design and RSA results.* ***(A)*** *Humans performed two tasks during a single EEG recording session. In the numerical decision task, participants viewed a stream of ten digits between 1 and 6, deciding whether the blue or orange numbers had the highest/lowest average. In the bandit task, participants learned about the reward probabilities of six images (bandits) and were asked to choose between two successive bandits to obtain a fixed reward. Numbers below each frame show ISI in ms.* ***(B)*** *RSA revealed a numerical and value distance effect from ~100 ms after stimulus onset. Inset shows magnitude model RDM. Shaded area represents SEM.* ***(C)*** *Averaged full RDM from 350-600 ms for numbers (1-6) and bandits (b1-b6). Upper left and lower right quadrants show representational similarity for numbers (N) and bandits (B) respectively, i.e. within-task RDM; lower left/upper right quadrants show cross-validated similarity between numbers and bandits (CV), i.e. between-task RDM.* ***(D)*** *Cross-temporal cross-validated RDM revealed a stable magnitude representation that was shared between the two tasks.* ***(E)*** *This effect was robust to the elimination of any single number/bandit pair. Correction for multiple comparisons was performed with cluster-based permutation tests,* $P_{cluster} < 0.005$.

learning in an initial learning phase. In the test phase, we asked participants to decide between two successive donkeys to obtain a reward, and estimated trial-wise subjective probability estimates for each bandit by fitting a delta-rule model to choices (*1*). Throughout these phases, the bandits were never associated with numbers in any way.

Next, we used RSA to examine the neural patterns evoked by bandits, ordered by their subjective reward probability. We found that multivariate EEG signals varied with subjective bandit ranks, with bandits that paid out with nearby probabilities eliciting more similar neural patterns (from ~100ms onwards; Fig. 1B, blue line). Our key question was whether there was a shared neural code for numerical magnitude and reward probability. We found that this was the case. For example, EEG signals elicited by digit "6" were more similar to those evoked by the most valuable bandit, and number 1 predicted the bandit least likely to pay out, with a similar convergence for intermediate numbers and bandits (Fig. 1C). Cross-validation of neural signals elicited by all numbers (1 to 6) and bandits (inverse ranks 1-6) was stable and reliable from 300-650 ms post-stimulus, as demonstrated by cross-temporal RSA (Fig. 1D) (*21*). This cross-validated pattern remained robust to the removal of any one of the six number/bandit pairs (Fig. 1E) or exclusion of the diagonal elements (Fig. S1A) and was more consistent within than between participants (Fig. S1B).

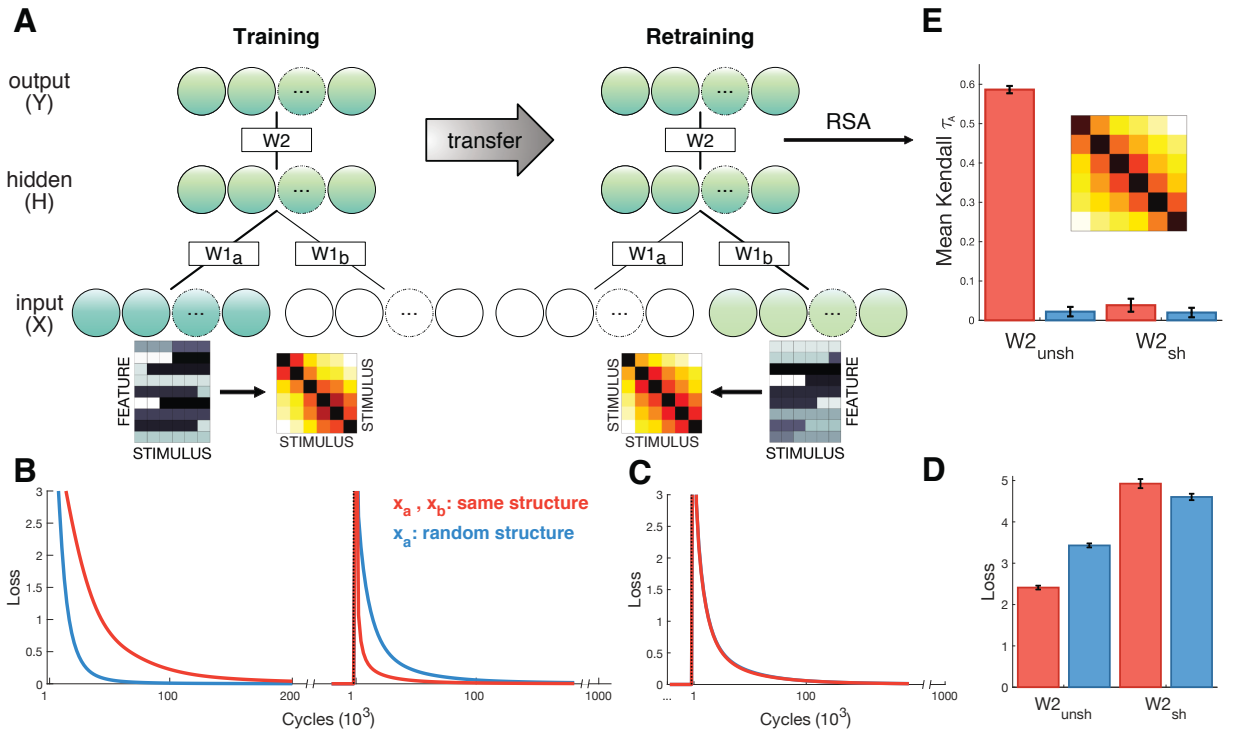

Fig. 2. *Neural network simulations.* ***(A)*** *Schematic depiction of network structure and training. The network was first trained to classify inputs* x_a *fed into units* X_A *(lower blue circles). Inputs* x_a *consisted of 6 stimuli that either exhibited gradual increasing dissimilarity or were shuffled as a control (stimulus RDMs shown next to examples of* x_a and x_b*). After convergence, the model was trained on new input* x_b *that were fed into a separate input stream* X_B *(lower green circles). Inputs* x_b *were different to* x_a *but exhibited the same similarity structure.* ***(B)*** *Loss plotted over the course of training (left panel) and retraining (right panel) for the test (red) and shuffled control (blue) conditions. Learning was faster during training for control stimuli, but retraining was faster when* x_a and x_b *exhibited shared similarity structure.* ***(C)*** *Loss for control simulations where W2 were shuffled between training and retraining, suggesting successful transfer depends on structure encoded in W2.* ***(D)*** *Mean loss for first 1000 cycles after retraining.* ***(E)*** *Cross-validation RSA on hidden unit activation for all stimuli in* x_a and x_b *after retraining. Hidden unit activations exhibit shared similarity structure only when W2 remains unshuffled and* x_a and x_b *share structure.*

To explore how these representations might be learned and generalized at the mechanistic level, we turned to a simple computational tool, a feedforward neural network (*22*). Unlike handcrafted probabilistic models, neural networks are not constrained to make inferences over structure, but structured representations may emerge naturally in the weights during training (*12*, *23*). Here, as a proof of concept, we confronted the network with two different stimulus sets in turn that (like our numbers and bandits) shared the same similarity structure. We then asked if the shared structure facilitates retraining on the second set after learning the first (Fig. 2A). The network was first trained on inputs x_1 arriving at input units X_A, and after convergence, retrained on inputs x_b fed into units X_B (where X_A and X_B are separate input modules that project to a common hidden layer H). Inputs x_b were 6 random vectors constructed to have the same continuously-varying similarity structure as the bandits, whereas inputs x_a consisted of either a different set of 6 random vectors with the same second-order structure, or a shuffled control. Relearning on x_b proceeded faster when inputs shared a common structure with x_a (Fig. 2B-D) and RSA conducted after retraining revealed reliable cross-validated patterns of activity in the hidden units for this condition alone (Fig. 2E), mirroring the result from the human neural data.

Our participants were all numerate adults with an intact sense of magnitude, and thus an equivalent "shuffled" control was unavailable for the human data. Nevertheless, building on past work that has described choice biases in numerical cognition and economic decisions (*24*), we asked whether individual differences in the mental number representation explained variance in the neural codes elicited by the bandits. We previously observed that participants overweighted larger magnitudes during the numerical decision task (*14*), e.g. numbers "5" and "6" had disproportionate impact on

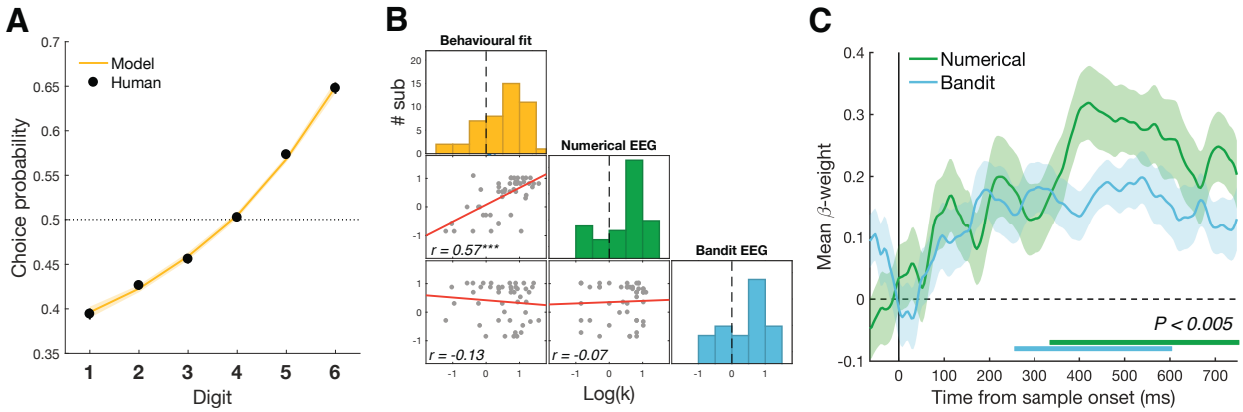

Fig. 3. *Psychometric and neurometric distortions.* ***(A)*** *Choice probabilities in the numerical task were modelled with a psychometric model that transformed input values into a decision value with exponent* k*. Participant's responses were best modelled by a distortion parameter* k > *1.* ***(B)*** *Distribution of estimated* k *for behavior in the numerical task and neural RDMs for both tasks. All measures were best described by* k > *1. Estimated distortions for behavior and EEG in the numerical task were highly correlated.* ***(C)*** *Choice distortions from the numerical task explained variance in the neural RDMs of the bandit task after partialling out choice behavior in the bandit task using competitive GLM ($P_{cluster} < 0.005$).*

averaging judgments, and this finding was replicated in the current data (Fig. 3A). Human choices were best fit by a power-law model in which participants averaged and compared distorted numerical values x^k with $k = 2.04 \pm 1.11$ ($k > 1$: $t(45) = 12.47$, $P < 0.001$). Turning to the neural data, we generated candidate representational dissimilarity matrices (RDMs) under the assumption that distance in neural space can like-wise be non-linear and best described by a distortion of the form x^k. We found that in both tasks the best fitting RDM was parameterized by $k > 1$ [numerical: $k = 1.73 \pm 0.82$, $t(45) = 14.34$, $P < 0.001$; bandit: $k = 1.72 \pm 0.85$, $t(45) = 13.65$, $P < 0.001$] (Fig. 3B). Moreover, estimated distortions for behavior and EEG in the numerical task were highly correlated, suggesting that they were driven by shared distortions in neural magnitude coding. To assess whether a distortion in the number line was also associated with choice behavior in the

bandit task, we predicted neural patterns in the bandit task from the choice distortions in the numerical task (after partialling out choice distortions estimated from the bandit task itself). Individual differences in warping of the number representation continued to explain variance in the neural bandit representation (Fig. 3C). This implies that humans used their intrinsic sense of magnitude when forming neural representations for the bandits in one-dimensional probability space.

Recent work has suggested that during categorization, posterior parietal neurons in the macaque monkey are strikingly low-dimensional, as if the parietal cortex were engaging in a gain control process that projected stimulus features or timings on a single axis (*25–27*). Indeed, focusing on centro-parietal electrodes, we observed a univariate positivity that varied with the magnitude of both numbers and reward probabilities (Fig. 4A-B). This signal resembles a previously described EEG signal, known as the CPP, that has been found to scale with the choice certainty in perceptual (*28*) and economic tasks (*29*). However, we note that in our numerical task the CPP followed an approximately ascending pattern from lower to higher numbers regardless of task framing or color category, and that the cross-validation effect persisted even after the CPP had been regressed out of the data (Fig. S2). This suggests that (a) the CPP in our task may represent a notion of magnitude, not a certainty signal; and (b) that this signal is not the sole driver of our multivariate findings. Nevertheless, to understand the dimensionality of the number and bandit representations (and the subspace in which they aligned), we used two dimensionality reduction techniques, singular value decomposition (SVD) and multidimensional scaling (MDS). First, using SVD, we systematically removed dimensions from the EEG data and recomputed our number-bandit cross-validation scores. We found that probabilistic reward learning was supported by a low-dimensional

neural magnitude code, with reliable effects persisting when all but 2 eigenvectors were removed

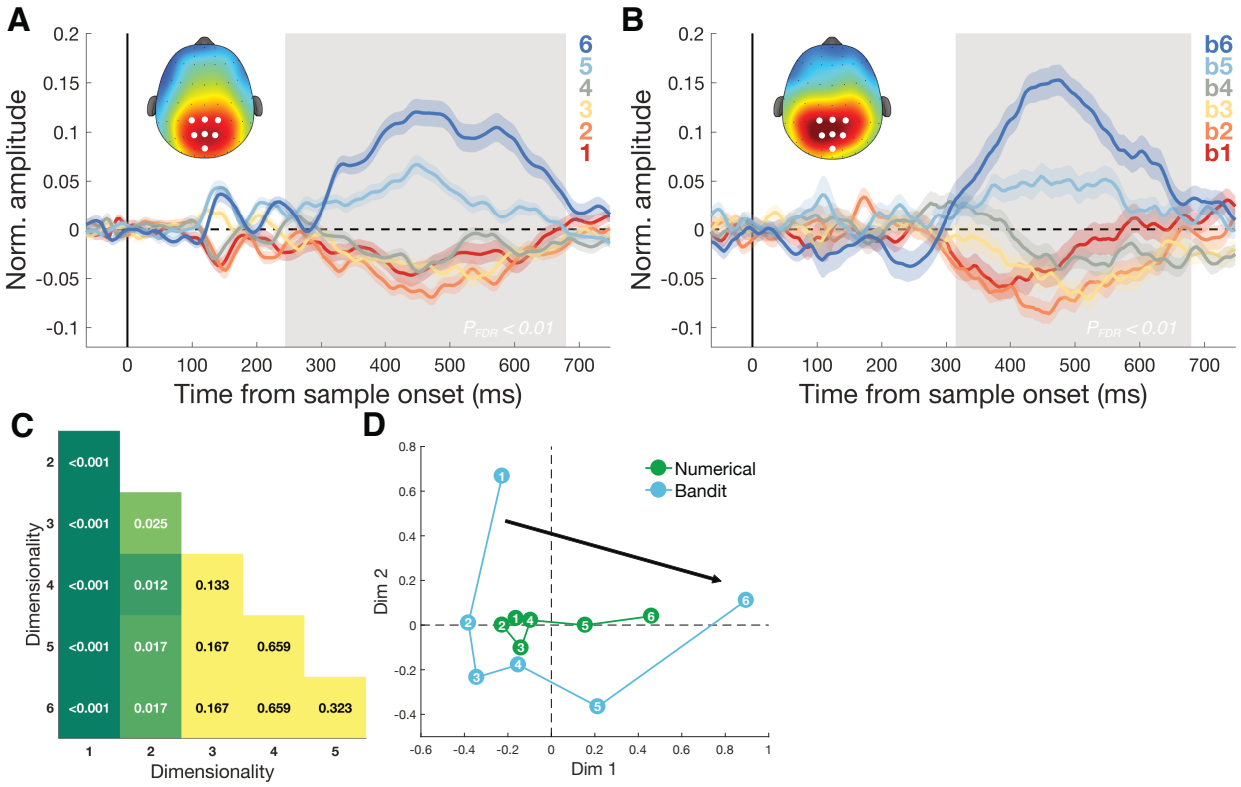

Fig. 4. *Dimensionality of magnitude representation.* ***(A)*** *Regression coefficients associated with numbers 1-6, independent of task framing (report highest/lowest) or category (blue/orange). Grey shaded area shows time of greatest disparity between signals (Kruskis-Wallis, $P_{FDR} < .01$). Scalp map inset shows response amplitude for digit 6 during identified time window. Colored shading represents SEM.* ***(B)*** *Equivalent analysis for bandits b1-b6 in the bandit task. Scalp map shows activation for highest subjectively valued bandit (b6).* ***(C)*** *Cross-validation after dimension reduction using SVD in each task. Each cell contains the p-value indicating the significance of the pairwise t-test comparing average cross-validation effect in the 350-600 ms time window under different dimensionalities of the data. Reduction to one dimension significantly reduced the size of the effect.* ***(D)*** *Multi-dimensional scaling (MDS) revealed two principal axes that describe the data: a magnitude axis approximately following the number/bandit order (along the black arrow) and a certainty axis distinguishing inlying (e.g. 3,4) from outlying (e.g. 1,6) numbers or bandits.*

from the data but attenuated when only a single dimension was retained in the EEG data. Statistical comparison suggested that within the 350-600 ms period for which significance was observed, the 2D and full (high-dimensional) solutions led to equivalent cross-validation but the reduction to 1D significantly reduced the effect (Fig. 4C). Secondly, we used MDS to visualize the first dimensions of the concatenated number/bandit data. This disclosed an axis pertaining to magnitude (Fig. 4D) and another for certainty along which, especially for the bandits, the large (or best) and small (or worst) items diverged from the others. In other words, the numbers and bandits align principally along a single magnitude axis but with an additional contribution from a second factor encoding choice certainty.

Together, these observations suggest that an abstract neural code for magnitude forms a "scaffold" for learning about the reward probabilities associated with novel stimuli. Rather than encoding stimulus value in an unstructured value function or lookup table (as is common in RL models), our data suggest that humans project available stimuli onto a one-dimensional axis that runs from "bad" to "good". This neural axis is aligned with the mental number line, suggesting that humans recycle an abstract concept of magnitude to encode reward probabilities. Learning a structured representation of value will have the benefit of allowing new inductive inferences, such as inferring transitive preferences among economic goods without exhaustive pairwise comparison, and facilitate read-out in downstream brain areas, related to the notion of a 'common currency' (*30*). Although the spatial resolution of EEG is limited, our univariate data point to the parietal cortex as one locus for the low-dimensional projection of value, where neural signals for magnitude have previously been proposed to provide a conceptual bridge between different metrics such as space, time and number (*16*, *17*, *31*, *32*).

References and Notes

1. R. S. Sutton, A. G. Barto, *Reinforcement learning : An introduction* (MIT press, Cambridge, MA, 1998).
2. R. J. Dolan, P. Dayan, Goals and habits in the brain. *Neuron*. **80**, 312–325 (2013).
3. W. Schultz, P. Dayan, P. R. Montague, A neural substrate of prediction and reward. *Science.* **275**, 1593–1599 (1997).
4. J. P. O'Doherty, P. Dayan, K. Friston, H. Critchley, R. J. Dolan, Temporal difference models and reward-related learning in the human brain. *Neuron*. **38**, 329–337 (2003).
5. V. Mnih *et al.*, Human-level control through deep reinforcement learning. *Nature*. **518**, 529–533 (2015).
6. B. M. Lake, T. D. Ullman, J. B. Tenenbaum, S. J. Gershman, Building Machines That Learn and Think Like People. *Behav. Brain Sci.* **40** (2017).
7. C. Kemp, J. B. Tenenbaum, The discovery of structural form. *Proc. Natl. Acad. Sci.* **105**, 10687–10692 (2008).
8. J. B. Tenenbaum, C. Kemp, T. L. Griffiths, N. D. Goodman, How to grow a mind: statistics, structure, and abstraction. *Science.* **331**, 1279–1285 (2011).
9. D. Gentner, Bootstrapping the mind: Analogical processes and symbol systems. *Cogn. Sci.* **34**, 752–775 (2010).
10. C. Kemp, J. B. Tenenbaum, S. Niyogi, T. L. Griffiths, A probabilistic model of theory formation. *Cognition*. **114**, 165–196 (2010).
11. D. Gentner, A theoretical framework for analogy. *Cogn. Sci.* **7**, 155–170 (1983).
12. J. L. McClelland *et al.*, Letting structure emerge: Connectionist and dynamical systems approaches to cognition. *Trends Cogn. Sci.* **14**, 348–356 (2010).

13. D. G. R. Tervo, J. B. Tenenbaum, S. J. Gershman, Toward the neural implementation of structure learning. *Curr. Opin. Neurobiol.* **37**, 99–105 (2016).

14. B. Spitzer, L. Waschke, C. Summerfield, Selective overweighting of larger magnitudes during noisy numerical comparison. *Nat. Hum. Behav.* **1**, 1–8 (2017).

15. A. L. Teichmann, T. Grootswagers, T. Carlson, A. N. Rich, Decoding Digits and Dice with Magnetoencephalography: Evidence for a Shared Representation of Magnitude. *J. Cogn. Neurosci.*, 1–12 (2018).

16. D. Bueti, V. Walsh, The parietal cortex and the representation of time, space, number and other magnitudes. *Philos. Trans. R. Soc. B*. **364**, 1831–1840 (2009).

17. V. Walsh, A theory of magnitude: Common cortical metrics of time, space and quantity. *Trends Cogn. Sci.* **7**, 483–488 (2003).

18. M. Fischer, S. Shaki, Number concepts: abstract and embodied. *Philos. Trans. R. Soc. B Biol. Sci.* **373**, 20170125 (2018).

19. N. Kriegeskorte, R. A. Kievit, Representational geometry: Integrating cognition, computation, and the brain. *Trends Cogn. Sci.* **17**, 401–412 (2013).

20. N. D. Daw, J. P. O'Doherty, P. Dayan, B. Seymour, R. J. Dolan, Cortical substrates for exploratory decisions in humans. *Nature*. **441**, 876–879 (2006).

21. J. R. King, S. Dehaene, Characterizing the dynamics of mental representations: The temporal generalization method. *Trends Cogn. Sci.* **18**, 203–210 (2014).

22. J. L. McClelland, T. T. Rogers, The parallel distributed processing approach to semantic cognition. *Nat. Rev. Neurosci.* **4**, 310–322 (2003).

23. A. M. Saxe, J. L. Mcclelland, "Exact solutions to the nonlinear dynamics of learning in deep linear neural networks", *arXiv preprint arXiv:1312.6120v3* (2013).

24. D. R. Schley, E. Peters, Assessing "Economic Value": Symbolic-Number Mappings Predict Risky and Riskless Valuations. *Psychol. Sci.* **25**, 753–761 (2014).

25. S. Ganguli *et al.*, One-Dimensional Dynamics of Attention and Decision Making in LIP. *Neuron*. **58**, 15–25 (2008).

26. J. K. Fitzgerald *et al.*, Biased Associative Representations in Parietal Cortex. *Neuron*. **77**, 180–191 (2013).

27. J. Wang, D. Narain, E. A. Hosseini, M. Jazayeri, Flexible timing by temporal scaling of cortical responses. *Nat. Neurosci.* **21**, 102–112 (2018).

28. R. G. O'Connell, P. M. Dockree, S. P. Kelly, A supramodal accumulation-to-bound signal that determines perceptual decisions in humans. *Nat. Neurosci.*, 1–7 (2012).

29. M. A. Pisauro, E. Fouragnan, C. Retzler, M. G. Philiastides, Neural correlates of evidence accumulation during value-based decisions revealed via simultaneous EEG-fMRI. *Nat. Commun.* **8**, 15808 (2017).

30. D. J. Levy, P. W. Glimcher, The root of all value: A neural common currency for choice. *Curr. Opin. Neurobiol.* **22**, 1027–1038 (2012).

31. M. V. Chafee, A Scalar Neural Code for Categories in Parietal Cortex: Representing Cognitive Variables as "More" or "Less." *Neuron*. **77**, 7–9 (2013).

32. C. Parkinson, S. Liu, T. Wheatley, A Common Cortical Metric for Spatial, Temporal, and Social Distance. *J. Neurosci.* **34**, 1979–1987 (2014).

33. D. H. Brainard, The psychophysics toolbox. Spat. Vis., 433–436 (1997).

34. M. Kleiner, D. Brainard, D. Pelli, "What's new in Psychtoolbox-3?" Percept. 3614 (2007).

35. A. Delorme, S. Makeig, EEGLAB: An open source toolbox for analysis of single-trial EEG dynamics including independent component analysis. *J. Neurosci. Methods*. **134**, 9–21 (2004).

36. H. Nili *et al.*, A Toolbox for Representational Similarity Analysis. *PLoS Comput. Biol.* **10** (2014).

37. E. Maris, R. Oostenveld, Nonparametric statistical testing of EEG- and MEG-data. *J. Neurosci. Methods*. **164**, 177–190 (2007).

38. Y. Benjamini, Y. Hochberg, Controlling the False Discovery Rate : A Practical and Powerful Approach to Multiple Testing, *J. R. Stat. Soc. Ser. B*. **57**, 289–300 (2009).

Acknowledgments: The authors would like to thank Zeb Kurth-Nelson for his insightful comments on an early draft of the manuscript and Mark Stokes for providing access to EEG equipment. **Funding:** This work was funded by an ERC Consolidator Grant to CS. FL is funded by the University of Oxford Clarendon Fund, the Department of Experimental Psychology, and a New College Graduate Studentship. BS was funded by DFG grant SP 1510/2-1. **Author contributions:** Conceptualization: FL, CS, BS; Methodology: FL, CS, BS, HN; Software: FL, BS, CS; Validation: FL; Formal Analysis: FL; Investigation: FL; Resources: CS; Data Curation: FL; Writing – Original Draft Preparation: CS, FL; Writing – Review & Editing: FL, CS, BS, HN; Visualization: FL; Supervision: CS, BS, HN; Project Administration: CS; Funding Acquisition: CS. **Competing interests:** Authors declare no competing interests. **Data and materials availability:** All data, code and materials to reproduce the analyses are available at https://github.com/summerfieldlab. Raw EEG data can be requested with the corresponding authors.