Value Gradient Rescales Grid-like Representations during Reward Learning

Hai-Tao Wu (wuhaitao@pku.edu.cn)

School of Psychological and Cognitive Sciences;

Beijing Key Laboratory of Behavior and Mental Health;

IDG/McGovern Institute for Brain Research;

Peking-Tsinghua Center for Life Sciences;

Peking University, Beijing 100871, China;

ORCID iD: https://orcid.org/0000-0003-3458-1459

Lusha Zhu (lushazhu@pku.edu.cn)

School of Psychological and Cognitive Sciences;

Beijing Key Laboratory of Behavior and Mental Health;

IDG/McGovern Institute for Brain Research;

Peking-Tsinghua Center for Life Sciences;

Peking University, Beijing 100871, China;

ORCID iD: https://orcid.org/0000-0001-8717-6356

Abstract

Cognitive maps are hypothesized to organize abstract features spatially, akin to physical environments, to guide decision and generalization. We propose that to better facilitate reward learning, these maps may distort by how fast value changes (i.e., value gradients), over-representing dimensions where feature changes yield larger reward differences. Using fMRI, we tested this during a reinforcement learning task with jellyfish images varying in two features (spot number, tentacle number) as options. Participants learned value maps where one feature had twice the reward sensitivity of the other. Under the assumption of internal maps scaled by value gradients, we observed six-fold periodic BOLD signals—a signature of grid-like coding—in entorhinal cortex (EC) and medial prefrontal cortex (mPFC). Compared with a range of alternative scales, signal strength peaked around value gradient scale and correlated with choice accuracy. These results point to a possibility that cognitive maps may be optimally constructed to reflect the reward structure in service of goal-directed behavior.

Keywords: cognitive map, grid-like coding, value gradient, fMRI

Introduction

Grid-like codes—periodic neural representations exhibiting six-fold symmetry—have been identified in the human entorhinal cortex and medial prefrontal cortex (mPFC) during virtual navigation tasks(Doeller, Barry & Burgess, 2010) and, more recently, abstract cognitive tasks (Constantinescu, O'Reilly & Behrens, 2016; Park, Miller & Boorman, 2021). Such geometric neural representation is widely hypothesized to reflect internal cognitive maps of task structure, functioning as a metric system for navigating both physical and conceptual spaces (Giocomo, Moser & Moser, 2011; Stachenfeld, Botvinic, & Gershman, 2017).

While early studies emphasized the role of grid-like codes in encoding the statistical structure of physical or conceptual spaces in a context-invariant manner (Fyhn et al., 2007; Hafting et al., 2005), emerging evidence suggests that these representations may be shaped by goal-directed processes. Specifically, rodent studies of spatial navigation demonstrate that grid patterns can be flexibly reshaped by goal locations, allowing the integration of reward information into the internal cognitive map (Boccara et al., 2019). This was explained as an over-representation of high value location, which is behaviorally important and thus frequently visited (Stachenfeld, Botvinic, & Gershman, 2017). Moreover, grid-like codes are identified during reward learning or value-based choices processes in humans and non-human primates(Park, Miller & Boorman, 2021; Bongioanni et al., 2021). These neural patterns are believed to represent feature space, which is important for predicting future reward in value learning (Gustafson, & Daw, 2011).

Building on prior findings, we explore a novel possibility that grid-like codes of abstract value map may be also sensitive to *reward gradient*—the direction of the steepest change in the value function (i.e., where small changes in a feature will yield largest reward differences). We reason that, to better facilitate reward learning, the internal representation of value space may be flexibly modulated by this gradient—stretching dimensions where value changes more rapidly to enhance representational precision. Of note, this hypothesis is not mutually exclusive with prior accounts emphasizing the over-representation of high-value locations. Rather, it is dissociable: a region associated with low reward magnitude but high local value gradient may still warrant enhanced resolution to support efficient learning and generalization.

Specifically, we predict that during reward learning, the cognitive map of value space is anisotropically stretched along feature dimensions in proportion to the local value gradient (Fig. 1A). As such distortions would alter the geometry of the underlying representational space, and thus the internal directions of 'navigation' through it, we propose that such modulation would affect the detection of sixfold periodicity of BOLD by directions: Six-fold periodicity in neural signals should be maximized when analysis is aligned with the geometry of the subject's internal cognitive map.

Design

We tested the hypothesis in a reward learning task with spatially correlated reward, generated by a value function unknown to subjects. Subjects need to choose between jellyfish images differing in spot count (X: 2-11) and tentacle count (Y: 2-11). Each jellyfish was assigned with a value, which is determined by its X and Y values. In particular, one dimension has twice the influence on value as the other, with different influential dimension in the two between-subject conditions (Figure A). Subjects performed a value learning (VL) task inside fMRI. On each trial, subjects chose between two jellyfishes, where the chosen value would later be delivered. The true value of both options were revealed after choices. Given the large number of possible stimuli (100 in total), subjects can hardly use a simple associative learning strategy but rather need to learn to underlying value function for generalizing value dependency to novel options. For each subject we derived their subjective relative importance of the two features from their choices, by which the dimensions of cognitive map are assumed to be stretched.

Results

Behavioral results indicated that participants quickly learned the value mappings, with choice accuracy rising to 90% by the second session (50 trials in 1 session; Figure B). Logistic regression analysis of choices (Choice ~ logit($\beta_x * \Delta X + \beta_y * \Delta Y$)) revealed that participants' decisions depended more strongly on the feature dimension with greater value influence(Figure C), critical to our further neural tests. Noticing most subjects did not actually learned the exact ground truth (relative importance 2:1), thus subjective value gradient would be assumed as scaling parameter for cognitive map.

fMRI analyses identified six-fold periodic signals in both EC and mPFC when presuming feature space scaled by subjective value gradients. Following previous studies, cross validation analysis (CV) of a six-fold periodicity across sessions (later sessions, 2-6), where training sessions was fit by a sixfold model to estimate the phase ρ of sixfold periodicity ($BOLD \sim cos(6 * (\theta - \rho))$; θ: navigation direction on each trial in gradient scaled space; ρ: individual parameter, phase of sixfold periodicity). The estimated ρ were then used to predict BOLD in the remaining one testing session. In a whole-brain CV analysis we identified lEC, rEC and mPFC as three ROIs showed potential grid like code where grid cells were previously recorded (Figure D). Further CV analysis in these ROIs demonstrated consistent sixfold effect across sessions (predictive lEC testing sessions using lEC training session: cross subjects mean spearman rho=.04, P =.001; rEC-rEC, mean rho= .04,P<.001; mPFC-mPFC, mean rho = .02; rEC-lEC, mean rho=.03, P<.001; mPFC-lEC, mean rho=.005, P=.65; mPFC-rEC, mean rho=.02, P=.03; Figure E). When navigation directions (θ-ρ) were binned into 30° intervals, all three regions of interest showed clear hexagonal modulation patterns(Figure F, G &H). Notably, the effect size in left EC correlated significantly with both choice accuracy (Pearson r = 0.38, p = 0.006; Figure L) and choice sensitivity to both features (r = 0.33, p = 0.018; Figure M), suggesting this sixfold effect in gradient scaled space is related to behavioral performance.

The specificity of these effects was confirmed through comparison with alternative scaling parameters. For each subjects, a range of alternative scaling parameters were assumed. Under these alternatively scaled geometry, the six-fold periodicity was strongest when using the subjective value-gradient-scaled geometry (Figure I, J &K), supporting our hypothesis that grid representations adapt to value gradients.

Our findings provide evidence for grid-like coding of value-gradient scaled feature spaces, with the strength of this representation relating to individual differences in learning performance. The results suggest that cognitive maps of abstract features are dynamically warped according to value gradients, potentially to optimize reward learning and generalization. Extend the concept of fitness maximization beyond perception fields to higher-level decision processes.

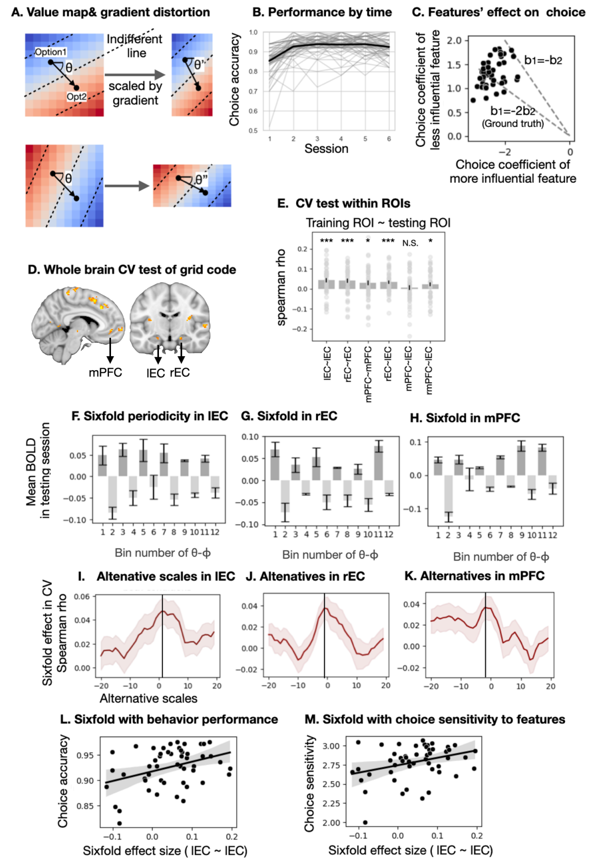

Figure A. left: Value map design(two conditions); right: predicted cognitive map scaled by value gradient. dashed line indicate the direction with identical value. When comparing the two options, subjects were hypothesized to mentally 'navigate' between them, where the navigation direction θ is affected by gradient scaling.

(B) Choice accuracy exceeded 90% after session 2.

(C) Regression coefficients from (Choice ~ logit(dx + dy)). Horizontal axis index the influential feature. Subjects showed larger dependency on the influential dimension (below the b1=-b2 line).

(D) Whole-brain analysis revealed sixfold periodicity in lEC/rEC/mPFC (p<0.005 uncorrected, k>10) in gradient-scaled geometry.

(E) ROI analysis confirmed sixfold periodicity (grid angles 2-6) and cross-ROI consistency (lEC-rEC/mPFC-rEC) in space scaled by value gradient.

(F-H) Sixfold modulation in 30° bins of $(\theta - \rho)$: (F) lEC, (G) rEC, (H) mPFC .

(I-K) Test of alternative scales . Zero point of X axis indicates each participants' estimated subjective value gradients. Tested scaling range of influential dimension: 1.1^{-20}~1.1^{20} (0.15~6.73) times. I: lEC; J: rEC; K: mPFC.

(L) Sixfold effect size in lEC of each subjects as estimated in Figure 5 positively correlates with their choice accuracy (r = 0.38, p = 0.006).

(M) Sixfold effect in lEC positively correlated with choice sensitivity to features (r = 0.33, p = 0.018). Choice sensitivity: rooted sum square of the two coefficients from the regression Choice ~ logit(dx + dy).

Acknowledgments

This work is supported by NSFC (T2421004, 32441104, 32325023 to L.Z.), STI2030-Major Projects (2022ZD0205100 to L.Z.) and CPSF (2022TQ0013 to Q.M.).

References

Boccara, C. N., Nardin, M., Stella, F., O'Neill, J., & Csicsvari, J. (2019). The entorhinal cognitive map is attracted to goals. *Science*, *363*(6434), 1443-1447.

Bongioanni, A., Folloni, D., Verhagen, L., Sallet, J., Klein-Flügge, M. C., & Rushworth, M. F. (2021). Activation and disruption of a neural mechanism for novel choice in monkeys. *Nature*, *591*(7849), 270-274.

Constantinescu, A. O., O'Reilly, J. X., & Behrens, T. E. (2016). Organizing conceptual knowledge in humans with a gridlike code. *Science*, *352*(6292), 1464-1468.

Doeller, C. F., Barry, C., & Burgess, N. (2010). Evidence for grid cells in a human memory network. *Nature*, *463*(7281), 657-661.

Fyhn, M., Hafting, T., Treves, A., Moser, M.B., and Moser, E.I. (2007). Hippo- campal remapping and grid realignment in entorhinal cortex. Nature *446*, 190–194.

Gustafson, N. J., & Daw, N. D. (2011). Grid cells, place cells, and geodesic generalization for spatial reinforcement learning. *PLoS computational biology*, *7*(10), e1002235.

Giocomo, L. M., Moser, M. B., & Moser, E. I. (2011). Computational models of grid cells. *Neuron*, *71*(4), 589-603.

Hafting, T., Fyhn, M., Molden, S., Moser, M.B., and Moser, E.I. (2005). Micro- structure of a spatial map in the entorhinal cortex. Nature *436*, 801–806.

Park, S. A., Miller, D. S., & Boorman, E. D. (2021). Inferences on a multidimensional social hierarchy use a grid-like code. *Nature neuroscience*, *24*(9), 1292-1301.

Schaffner, J., Bao, S. D., Tobler, P. N., Hare, T. A., & Polania, R. (2023). Sensory perception relies on fitness-maximizing codes. *Nature Human Behaviour*, *7*(7), 1135-1151.

Stachenfeld, K. L., Botvinick, M. M., & Gershman, S. J. (2017). The hippocampus as a predictive map. *Nature neuroscience*, *20*(11), 1643-1653.