RewCon Modeling: Second Attempt

Takao Noguchi
t.noguchi@ucl.ac.uk

The rational model of categorization (RMC; Anderson, 1991) was fitted to participants' responses during the training phase. In particular, we used the local MAP version of the RMC as described by Anderson (1991). The modeling exercise is summarized in this document. First, I briefly describe the RMC, with the aim to introduce notations I use to describe model-based estimates. Please note that the description below follows the convention by Sanborn, Griffiths, and Navarro (2010), not by Anderson (1991).

Model — Rational model of categorization (RMC)

Suppose a learner has observed $n-1$ cue-combinations $\{x_1, x_2, \ldots, x_{n-1}\}$ with corresponding category labels $\{y_1, y_2, \ldots, y_{n-1}\}$. In the RewCon experiment, x_i is a pair of cues presented in the ith trial, and y_i is a corresponding category (hat or glove). In the RMC, each of these cue-combinations fits into a cluster. The cluster label for the ith cue-combination is denoted as z_i.

Probability of response

Following Davis, Love, and Preston (2012), we introduced a scaling parameter γ to the RMC, in computing the probability of judging x_n as Category w:

$$p^*(y_n = w \mid x_n) = \frac{p(y_n = w \mid x_n)^\gamma}{\sum_u p(y_n = u \mid x_n)^\gamma.} \tag{1}$$

Here, $p(y_n = w \mid x_n)$ is what the RMC predicts as the probability that the nth cue-combination fits into category w, but we used $p^*(y_n = w \mid x_n)$ to predict participants' responses.

Last updated: January 17, 2017.

Drawing an inference with the RMC

Then,

$$
\begin{aligned}
p(y_n = w \mid x_n) &= \sum_{k \in \mathbb{Z}} p(z_n = k \mid x_n)\, p(y_n = w \mid z_n = k) \\
&= \sum_{k \in \mathbb{Z}} \frac{p(z_n = k)\, p(x_n \mid z_n = k)}{p(x_n)}\, p(y_n = w \mid z_n = k) \\
&= \sum_{k \in \mathbb{Z}} \frac{p(z_n = k)\, p(x_n \mid z_n = k)}{\sum_{s \in \mathbb{Z}}\, p(z_n = s)\, p(x_n \mid z_n = s)}\, p(y_n = w \mid z_n = k).
\end{aligned} \tag{2}
$$

Here, $\mathbb{Z}$ is a set of all the possible clusters to which the nth cue-combination can be assigned. The three terms in Equation 2 are described below in turn.

First, the probability that the nth cue-combination fits into a cluster k is given by:

$$
p(z_n = k) = \begin{cases} \dfrac{c\, m_k}{(1-c) + c\,(n-1)} & \text{if} \quad m_k > 0 \\[2ex] \dfrac{(1-c)}{(1-c) + c\,(n-1)} & \text{if} \quad m_k = 0 \end{cases} \tag{3}
$$

where c is a parameter called the coupling probability and m_k is the number of cue-combinations assigned into cluster k.

In the RewCon experiment, two cues were independent. Therefore,

$$
p(x_n \mid z_n = k) = \prod_{d \in D} p(x_{n,d} \mid z_n = k), \tag{4}
$$

where D is a set of first and second cues. The above term is computed with

$$
p(x_{n,d} = v \mid z = k) = \frac{B_{v,d} + \beta_c}{m_k + J_d\, \beta_c}, \tag{5}
$$

where $B_{v,d}$ is the number of cue-combinations in cluster k with dth cue being v, and J_d is the number of possible cues for the dth cue (four in the RewCon).

Similarly, the probability that the nth cue-combination has category label w, given a cluster, is given by:

$$
p(y_n = w \mid z = k) = \frac{B_w + \beta_l}{B_{\cdot} + J\, \beta_l}, \tag{6}
$$

where B_w is the number of observed cue-combinations with category label w in cluster k, $B_{\cdot}$ is the number of cue-combinations in cluster k, and J is the number of possible category labels (two in RewCon).

Learning with the RMC

The learning in the RMC is to assign an cue-combination into a cluster. This cluster assignment is performed with the probability of each cluster. Specifically, the nth cue-combination is assigned to a cluster with the largest probability:

$$
\begin{aligned}
z_n &= \arg\max_k \; p(z_n = k \mid x_n, y_n) \\
&= \arg\max_k \; p(z_n = k)\, p(x_n \mid z_n = k)\, p(y_n \mid z_n = k).
\end{aligned} \tag{7}
$$

This is computed with Equations 3, 4 and 6.

Parameter Estimation and Model Fit

The above model, when applied to the RewCon experiment, has three parameters: scaling γ, coupling probability c, sensitivity parameter for a cue β_c, and sensitivity parameter for category label β_l.

We fixed the values of the sensitivity parameters: $\beta_c = 0.01$ and $\beta_l = 1.00$. Then I identified a value range of coupling probability, which produces the reasonable number of clusters (i.e., not hundreds) for the rewcon experiment. With the range between 0.17 and 1.00, the RMC produces 16 or less clusters, and thus, the value of coupling probability is restricted within this range.

Then, I ran two iterations of two estimation: one where γ is fixed at 1.0 (`fixed_gamma`) and another where γ is not fixed but shared across participants (`free_gamma`). The model was fitted to participants' behavioral responses, and the parameter values were estimated with the maximum likelihood criterion. The likelihood of the estimated parameters is -4294.21 (AIC: 8614.42) for `fixed_gamma` and -4285.68 (AIC: 8598.36) for `free_gamma`.

The estimated parameter values are saved in "`estimated_parameter.csv`". This csv file also includes the number of clusters created after the 288 training trials.

Model-based Estimates

Trial-by-trial estimate

Trial-by-trial estimates are recorded in "`trial_by_trial_estimate.csv`" along with the trial-by-trial data from the experiment.

- `p_correct`

  Probability of making a correct response, based on a pair of cues. In the above notation, this is $p^*(y_n = \text{correct label} \mid x_n)$. A related estimate ($1-$ `p_correct`) was used by Davis et al. (2012).

- `p_new_cluster`

  Probability of creating a new cluster, based on a pair of cues.

- `recognition_strength`

  Recognition strength, $p(x_i)$, before a cue-combination is assigned to a cluster. This was used by Davis et al. (2012). This estimate relates to the extent to which an cue-combination matches the RMC's stored category representations.

- `cluster_entropy`

  The entropy of $p(z_i \mid x_i)$. This estimate indexes the extent to which the RMC is uncertain about which cluster a cue-combination belongs to. This was used by Davis et al. (2012).

Note

Difference between this attempt and the previous attempt

Previously, we tried the particle filter version of the RMC (as proposed by Sanborn et al., 2010), while in this second attempt, we are using the local MAP version of the RMC (as proposed by Anderson, 1991). The difference between the two versions is in how a cue-combination is assigned to a cluster.

With the particle filter version, the nth cue-combination is assigned to a cluster with the probability $p(z_n = k \mid x_n, y_n)$. Therefore if this probability is 0.1, there is .1 probability that the cue-combination is assigned to this cluster. This probabilistic assignment leads to stochasticity in learning: even with the same parameter values, the RMC can assign the cue-combination to a different cluster. To manage this stochasticity, we had to simulate the learning 2,000 times and average out the estimates.

With the local MAP version, the cluster assignment is deterministic. The nth cue-combination is assigned to a cluster with the largest probability $p(z_n = k \mid x_n, y_n)$. Thus with the same parameter values, the RMC assigns the cue-combination to the same cluster. Consequently, we did not have to simulate the model many times.

In addition, as we restricted the value range for coupling probability, the number of clusters does not grow to hundreds, and the estimates do not have much between-block variances.

Why 5 clusters, instead of 4 clusters

The estimates suggest that after all the 288 trials, the RMC produced 1, 2, 5 or 16 clusters. A 5-cluster solution may appear odd, as it does not clearly map onto the category structure.

A closer look tells us that a 5-cluster solution is essentially a 4-cluster solution. When 5 clusters are created, one of them has only one cue-combination in it. What exactly is this cue-combination appears to depend on a presentation order.

Take, as an example, estimates for Participant 1. Cluster 4 contains only one instance of Cue-combination 8-4. Other instances of 8-4 appear to have fit into Cluster 2. Here, Cluster 2 hosts 4 cue-combinations: 1-3, 8-3, 1-4, and 8-4. To understand this, we take a look at the trial order below. Relevant cue-combinations are indicated with bold font.

Trial	Cue 1	Cue 2	Content of Cluster 2
1	7	4	—
2	**1**	**3**	**1-3**
3	1	2	1-3
4	**8**	**4**	**1-3**
5	7	5	1-3
6	7	2	1-3
7	**8**	**3**	**1-3, 8-3**
8	1	5	1-3, 8-3
9	6	5	1-3, 8-3
10	8	2	1-3, 8-3
11	7	3	1-3, 8-3
12	8	5	1-3, 8-3
13	**1**	**4**	**1-3, 8-3, 1-4**
14	6	4	1-3, 8-3, 1-4

Among the four cue-combinations, 1-3 was the earliest one to be presented at Trial 2. Then 8-4 was presented at Trial 4. At Trial 4, however, Cluster 2 hosted only 1-3 and had nothing in common with 8-4. As a result, a new cluster was created to host 8-4 at Trial 4.

Then 8-3 was presented at Trial 7, and 1-4 was presented at Trial 13. Both got assigned to Cluster 2. So after Trial 13, Cluster 2 hosted 1-3, 8-3, and 1-4, and so, Cluster 2 contained Cues 8 and 4 — the identical values to Cue-combination 8-4.

In addition, Cluster 2 had more cue-combinations than Cluster 4 (3 vs 1 after Trial 13). As the RMC is more likely to assign a cue-combination to a cluster which hosts more cue-combinations, 8-4 subsequently got assigned to Cluster 2 and not to Cluster 4, effectively abandoning Cluster 4. This process, creation and abandonment of Cluster 4, led to a 5 cluster solution, rather than a 4 cluster solution.

As the trial order does not appear to have been randomized (I may be wrong here), the same explanation applies to other participants where the RMC produced 5 clusters.

Differences between the 1- and 2-cluster solutions

The 1-cluster solution is estimated for some participants, while the 2-cluster solution is estimated for some others. As both 1- and 2-cluster solutions predict 0.5 overall accuracy, this difference in the estimations may appear strange.

In fact, the 1- and 2-cluster solutions make slightly different accuracy predictions especially during early trials. To illustrate, let us look at the example below.

Trial	Cue 1	Cue 2	Category
1	7	4	hat
2	1	3	glove
3	1	2	hat
4	8	4	glove
5	7	5	glove
6	7	2	glove
7	8	3	glove
8	1	5	hat
9	6	5	glove
10	8	2	hat

The 1-cluster solution puts all the cue-combinations into one cluster, and its categorization is primarily driven by the number of category instances. During Trials 1 to 8, for example, 5 out of 8 cue-combination fit into the glove category, so at Trial 9, probability of the glove category is predicted to be about $\frac{5}{8}$. Here, both Cue 1 and 2 are completely ignored.

In contrast, the 2-cluster solution pays attention to cues. At Trial 9, a new cue is presented as Cue 1, so based on Cue 1 alone, both category is equally likely. Cue 2 matches previous two instances (Trials 5 and 8), one of which fits into the glove category, so based on Cue 2 alone, both category is equally likely as well. Then, the probability of the glove category at Trial 9 is predicted to be about $\frac{1}{2}$.

Therefore, if participant responded “glove” at Trial 9, the 1-cluster solution is considered more likely than the 2-cluster solution.

Overall accuracy and the number of clusters

Figures 1 and 2 summarize relations between overall accuracy and the number of clusters. Here, overall accuracy is the proportion of correct judgments during the 288 training trials, and the number of clusters is counted after the 288th trial. Participants who did not appear to have learned much got 1 or 2 cluster solutions. Participants who got near-perfect accuracy quickly got 5 cluster solutions, and other participants who slowly got near-perfect got 16 cluster solutions.

Block accuracy

Figures 3 and 4 summarize by-block average accuracy for each participant.

References

Anderson, J. R. (1991). The adaptive nature of human categorization. *Psychological Review*, *98*, 409–429.

Davis, T., Love, B. C., & Preston, A. R. (2012). Striatal and hippocampal entropy and recognition signals in category learning: simultaneous processes revealed by model-based fMRI. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, *38*, 821–839.

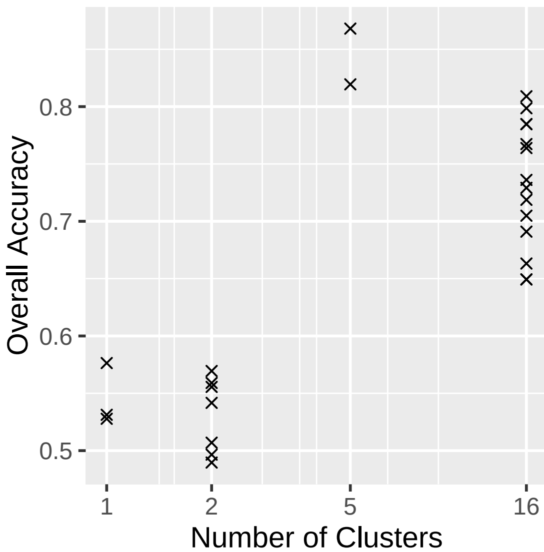

Figure 1. Overall accuracy of judgment and the number of clusters, when γ is fixed at 1.0.

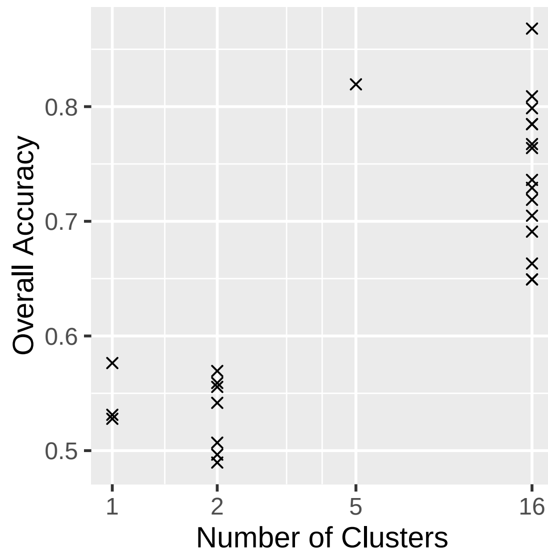

Figure 2. Overall accuracy of judgment and the number of clusters, when γ is not fixed.

Sanborn, A. N., Griffiths, T. L., & Navarro, D. J. (2010). Rational approximations to rational models: alternative algorithms for category learning. *Psychological Review*, *117*, 1144–1167.

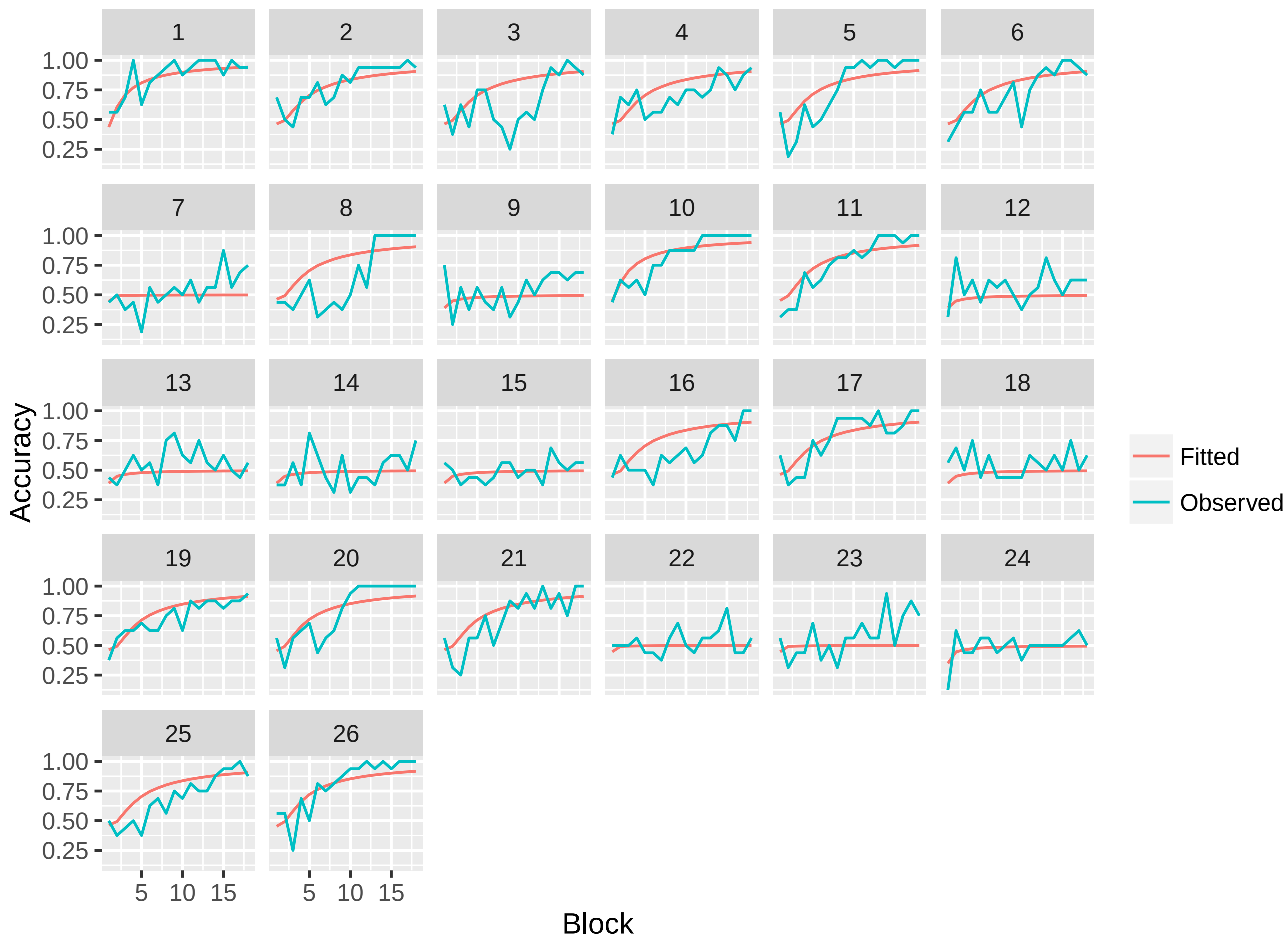

Figure 3. Average accuracy per block of judgment, when γ is fixed at 1.0.

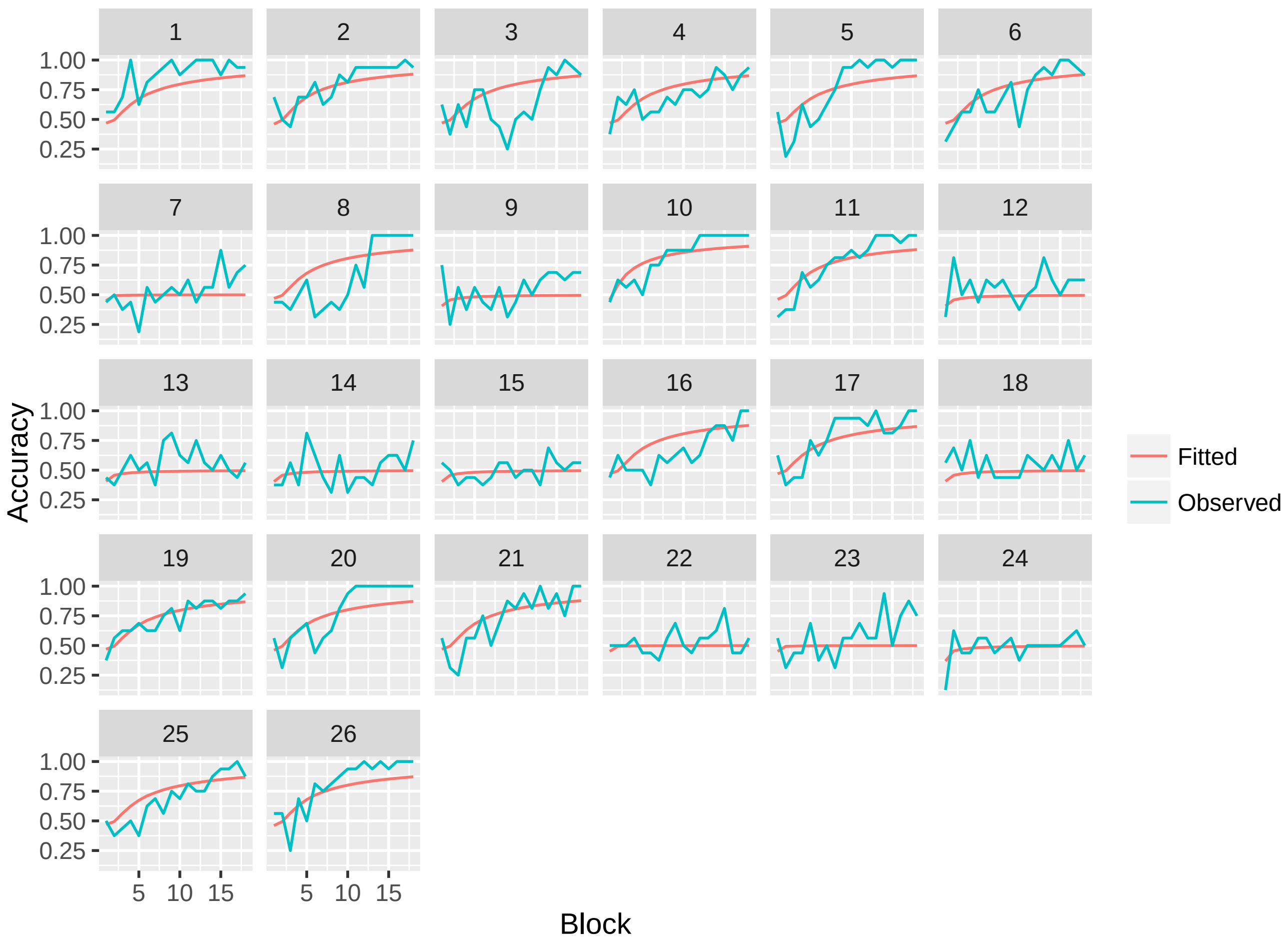

Figure 4. Average accuracy per block of judgment, when γ is not fixed.