Cognitive Symmetries as Bases for Anticipation: a Model of Vygotskyan Development Applied to Word Learning

Tatsuji Takahashi*, Shuji Shinohara**, Kuratomo Oyo* and Asaki Nishikawa***

**School of Science and Engineering, Tokyo Denki University, Hatoyama, Hiki, Saitama, 350-0394, Japan. +81492966403 - tatsujitakahashi@me.com

**Advanced Algorithms & Systems, Co. Ltd.

***Flucto-Order Functions Research Team, RIKEN-HYU Collaboration Research Center, RIKEN Advanced Science Institute

Abstract In studying human cognition, it is now broadly approved that the study of cognitive biases is indispensable. Among many cognitive biases proposed, we focus on two symmetrical biases: symmetry and mutual exclusivity bias. Implementing the two biases in a probabilistic framework on covariation information, we test our loosely symmetric (LS) model in word learning tasks, in comparison with the ordinary conditional probability and the totally symmetric/biased probability. LS is shown to break a trade-off in the three tasks. It is argued that LS is a model of development in the sense of Vygotsky, where top-down/deductive and bottom-up/inductive processes crisscross.

Keywords : cognitive bias; heuristics; stimulus equivalence; mutual exclusivity

1 Introduction

Human intuition is biased. Human inference is not purely logical. It is extensively argued in psychology. The illogicality is conceptualized as biases and heuristics [2]. Most of them are considered as the default modes of cognition for rational information processing, procured through the course of evolution pathways. One of the problems of studying the biases is the lack of organization on the huge variety of biases and heuristics. We presume the key is their symmetry and anticipatory nature of being in the world. In this article, we study two symmetrical biases in relation to a trade-off between learning and communication in word learning. We test an adjustment mechanism of the biases, LS (loosely symmetric) model ([11], [12]) discovered by Shinohara [8], and show that it breaks the trade-off. We also argue that it is a model embodying the developmental theory by Lev Semyonovich Vygotsky, where top-down/deductive and bottom-up/inductive processes intersect. First we argue that the symmetrical biases form the bases for anticipation that is essential to cognition in general.

International Journal of Computing Anticipatory Systems, Volume 24, 2010
Edited by D. M. Dubois, CHAOS, Liège, Belgium, ISSN 1373-5411 ISBN 2-930396-12-1

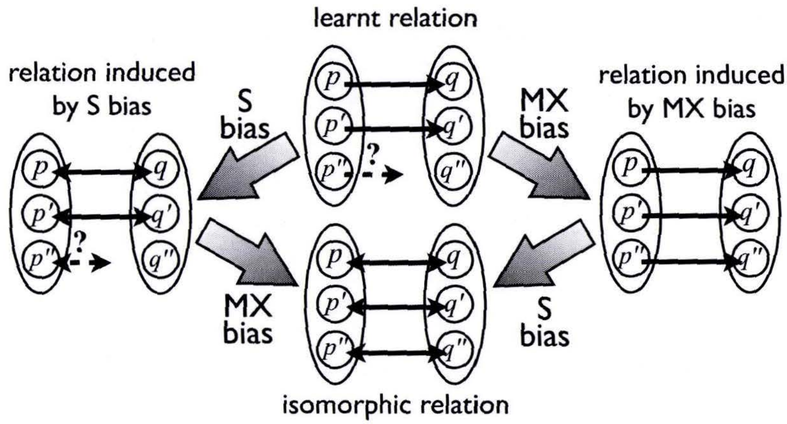

Fig. 1: The effect of S and MX biases in forming relations from a learnt one.

2 The Cognitive Symmetries and Anticipation

The two symmetrical biases we study in this paper have appeared originally in the study of word learning in a broad sense. One is mutual exclusivity (MX) bias [6]. Given a binary relation pRq between a stimulus p and q learnt, MX bias is to believe $\bar{p}R\bar{q}$, where $\bar{p} \neq p$ and $\bar{q} \neq q$. The other one is symmetry (S) bias [9]. It is to believe qRp from learning pRq. See Fig. 1.

S bias has been formulated in the study of "stimulus equivalence" in the behavior analysis or comparative psychology (Sidman [9]) and of "biconditionality" in causal induction (Hattori [3]) and other higher-level cognitive tasks [5]. It is strongly related to anticipation, at least via abduction. The bias can provide a basis for internalist anticipation. When you internally anticipate an event B, acting toward it, you need to abduct an action A about which it is known that A induces B, a directed relationship $A \to B$. By doing A, you try to create the future event B and manage to meet the anticipated B. In the inference it can take the form of deduction if you throw $A \to B$ in reverse. Deduction is to infer consequent B from A and $A \to B$. Abduction is to infer antecedent A from $A \to B$ and B. It is not logically valid but human beings reason in that way, regardless of age, sex, race or capability [14].

MX bias has been proposed in the study of children's vocabulary learning (Markman [6]) in developmental psychology. Imagine that there are a familiar object with a known name and another novel object with its name unknown. When an infant hears a name previously unheard, she tends to consider that it is the name of the novel object. This tendency is called mutual exclusivity (MX) bias. It helps her to identify the referenced object in the absence of other cues such as her mother gesturing toward it. Formally speaking, this tendency is to believe $\bar{p} \to \bar{q}$ when $p \to q$ is accepted, where p, q are some stimuli and $\bar{p}$ is a stimulus different from p.

It is interesting that animals other than human only quite rarely show the symmetries [14]. It means that in a sense animals are more logical than us human. Actually, we laboriously learn to infer asymmetrically. It it would not be like this, learning mathematical logic must have been much easier, while we often confuse directed and asymmetrical relations between, e.g., necessary and sufficient conditions.

2.1 Symmetric Metaphysics: Relation as Object and Origin of Isomorphism

The S and MX biases provide the bases for anticipation in cognition: time reversing, causality inversion, future cause bringing about present effect, etc. It is these biases that make anticipation familiar, rather intrinsic, to human.

It is interesting to see the formal or metaphysical relationship between S and MX biases, that can also be seen to lead to logical contrapositives, $q \to p$ and $\neg p \to \neg q$. S makes relationships among objects atemporal and/or undirected, while MX renders relationships non-intersecting. Atemporal and independent mode of existence is what characterizes metaphysical objects (S). The relationships are treated as objects, if object is naively defined by a closed boundary of its own (MX). S and MX are, respectively, closely related to the existence of inverse map and one-to-one correspondence that are both structure-preserving conditions that lead a map to be isomorphic. S and MX biases lead to isomorphism when composed, as in Fig. 1. We now have a way to inquire into the human origin of the equivalence of these conditions for mathematical identity: isomorphism. It is a fundamental concept for human cognition, such as in one-to-one correspondence in counting.

Objectification of relation leads to a metaphysics in which objects and relations are not asymmetric, while usually they are rigorously distinguished because of the difference in logical type. In this way, the symmetrical cognitive biases are intimately related to the problems around the difference in logical type, such as self-reference. It would be interesting if the biases are what make human able to treat self-reference as the source of creativity and not just as logical contradiction or infinite oscillation between true and false [10].

2.2 Anticipatory Causality

Given a subjective sense of causation that p causes q, S bias subjectively identifies or confuses an effect q in the distant future, locating at the posterior point in the causal sequence, and a cause p that is at present realizable. It is that the future is at present. This brings about the sense of decidability of the future. One can choose p to cause q since they are equivalent. S bias is the possibility of anticipatory act. On the other hand, from p causing q, MX bias makes the subjective causal sequence mutually exclusive. It makes one think that when the anticipated q is invoked by the present option p, q can not be brought about by any other option $\bar{p} \neq p$. One must choose p to cause q. MX bias is the necessity of anticipatory action. In this

Table 1: A 2 x 2 contingency table for causal induction.

		posterior event	
		q	$\bar{q}$
prior event	p	a	b
	$\bar{p}$	c	d

Table 2: A totally symmetric 2 x 2 contingency table.

		posterior event	
		q	$\bar{q}$
prior event	p	$a+d$	$b+c$
	$\bar{p}$	$c+b$	$d+a$

sense, these biases form the bases for anticipation: *interchangeability of cause and effect* and *future effect bringing about present or past cause.*

3 The Probabilistic Representation

The MX and S biases both represent our intuition in causal induction. Let there be two events or stimuli p and q. When do we say that "p causes q"? One condition is high conditional probability of q given p, $P(q|p)$. It is enough for prediction. However, this condition is not sufficient for our intuitive feeling of causation. For example, let us always observe crows caw (p) before the sun rises (q). It means the conditional probability $P(q|p)$ is very high. However, we do not think the sunrise is caused by the crows cawing.

There are two conditions for us to feel causality [8]. One is that q does not occur when p does not. In probability it is expressed by simultaneously high $P(\bar{q}|\bar{p})$ with high $P(q|p)$. It is called a principle of causality. The other one is that p occurs when q does. This means high $P(p|q)$. Of course, the usual definition of probability does not satisfy them. So we denote a form of subjective causality estimation as $B, 0 \leq B \leq 1$. We define these two conditions as the expression of MX and S biases as follows:

$$B(q|p) = B(\bar{q}|\bar{p}) \qquad \text{(MX bias)} \qquad (1)$$

$$B(q|p) = B(p|q). \qquad \text{(S bias)} \qquad (2)$$

Another condition, law of excluded middle (XM), is logically important since it means the well-definedness of negation.

$$B(q|p) = 1 - B(\bar{q}|p). \qquad \text{(XM)} \qquad (3)$$

The models for causal induction studied in this article are NS (nonsymmetric), RS (rigidly symmetric) and LS (loosely symmetric). They are defined on a 2×2 contingency table as in Table 1. It represents covariation information between two events, p and q. NS is defined as the ordinary conditional probability.

$$NS(q|p) = a/(a+b) \qquad (4)$$

NS does not satisfy neither S nor MX, but it has XM.

By making the cells in Table 1 symmetric, we get RS from NS. Because $NS(p|q) = a/(a+c)$, the condition $b = c$ induces the complete S bias, $NS(q|p) = NS(p|q)$. MX bias is rigidly formed if $a = d$ holds additionally, since $NS(\bar{q}|\bar{p}) = d/(d+c)$. The equations $b = c$ and $a = d$ are satisfied in Table 2. The symmetries in the table are expressed in an equation as follows:

$$RS(q|p) = (a+d)/((a+d)+(b+c)) \tag{5}$$

Thus we get RS. RS has all of S, MX, and XM.

3.1 A Loosely Symmetric (LS) Model

The LS model featured in this article is defined as follows:

$$LS(q|p) = \frac{a + P(p|\bar{q})d}{a + P(p|\bar{q})d + b + P(p|q)c} = \frac{a + \frac{b}{b+d}d}{a + \frac{b}{b+d}d + b + \frac{a}{a+c}c}. \tag{6}$$

Inductive and deductive derivations of the LS model are respectively found in [11] and in [12]. Here we just mention that LS can be understood in two ways. On one hand, it is an NS-like formula $(a/(a+b))$ with the added extra terms $(b/(b+d))d$ to a and $(a/(a+c))c$ to b. On the other hand, LS can be understood to be kin to RS, with the term c and d weakened by the coefficients $a/(a+c)$ and $b/(b+d)$, respectively. The duality is the key for deriving LS. LS has XM and loosely satisfied S and MX, varying the intensity of the biases according to the situation, i.e., the value of (a, b, c, d).

4 Word Learning Tasks

The cognitive symmetries were proposed in the study of word learning in general. We apply the LS model to word learning tasks back. There is an infant and an adult. The adult knows the correspondence between labels and objects correctly, and she teaches them to the infant. The infant guesses the connection between labels and objects with a probabilistic model B. The learning is tested with three tasks. The infant demonstrates the correctness of her learning to the adult (task 1 and 2), and he also asks the adult for passing an object to him, by telling the label of the object (task 3). The former tasks test the correct reproduction of the memory. The latter more or less requires S bias, $B(o|l) = B(l|o)$, if not rigorously equal, where o is an object and l is a label.

We apply the causal models to the word learning tasks. For calculating $B(o_j|l_i)$, a correspondence from a label l_i to an object o_j, the cells in the contingency table, a, b, c and d, are as follows[1] (note that $\bar{o}$ ($\bar{l}$) means objects (labels) other than o (l).):

[1]This gives a generalization of LS to $m \times n$ contingency table other than in [12].

	o_1	$\cdots$	o_{j-1}	o_j	o_{j+1}	$\cdots$	o_m
l_1	x_{11}	$\cdots$	$x_{1(j-1)}$	x_{1j}	$x_{1(j+1)}$	$\cdots$	x_{1m}
$\vdots$	$\vdots$	d	$\vdots$	$\vdots$ c	$\vdots$	d	$\vdots$
l_{i-1}	$x_{(i-1)1}$	$\cdots$	$x_{(i-1)(j-1)}$	$x_{(i-1)j}$	$x_{(i-1)(j+1)}$	$\cdots$	$x_{(i-1)m}$
l_i	x_{i1}	$\cdots$ b	$x_{i(j-1)}$	x_{ij} a	$x_{i(j+1)}$	$\cdots$ b	x_{im}
l_{i+1}	$x_{(i+1)1}$	$\cdots$	$x_{(i+1)(j-1)}$	$x_{(i+1)j}$	$x_{(i+1)1(j+1)}$	$\cdots$	$x_{(i+1)m}$
$\vdots$	$\vdots$	d	$\vdots$	$\vdots$ c	$\vdots$	d	$\vdots$
l_n	x_{n1}	$\cdots$	$x_{n(j-1)}$	x_{nj}	$x_{n(j+1)}$	$\cdots$	x_{nm}

Fig. 2: The contingency table for word learning with the focus on an object o_j given a label l_i, for the estimation of correspondence from l_i to o_j with $B(o_j|l_i)$.

$$a = P(l, o) \tag{7}$$
$$b = P(\bar{l}, o) = P(o) - P(l, o) = \Sigma_{l' \neq l} P(l', o) \tag{8}$$
$$c = P(l, \bar{o}) = P(l) - P(l, o) = \Sigma_{o' \neq o} P(l, o') \tag{9}$$
$$d = P(\bar{l}, \bar{o}) = 1 - (a + b + c) = \Sigma_{l' \neq l} \Sigma_{o' \neq o} P(l', o') \tag{10}$$

When the inverse probability $B(o_j|l_i)$ (the probability of the inverse direction) is to be calculated, b and c in Fig. 2 are interchanged.

4.1 The Common Settings of Simulation

Here we give the three models (NS, RS and LS) the following three word learning tasks. There are objects and labels. Labels are attributes such as color and shape.

1. (What-task) Answer a label of a given object.
2. (Which-task) Answer an object that has a given label.
3. (Hand-me-task) Ask an adult to pass an object of desire by telling a label.

The common setting throughout the three tasks are as follows:

- Each object has two kinds of label (attribute): color and shape. The disjoint sets of colors and shapes are denoted by C and S. The set of labels is $L = C \bigcup S$. It is assumed that there are more shapes than colors ($|C| < |S|$). One reason comes from our discernment and language that distinguishes more shapes than colors. In addition, formally speaking, infinite shapes can be recursively generated and recognized but colors can not.
- The set of objects, O, is defined to be the product of colors and shapes, $O = C \times S$, hence $|O| = |C| \times |S|$. There are as much objects as the combination of colors and shapes.

In Table 3, there is an example of the joint occurrence rates between two attributes, $|C| = 2$ colors and $|S| = 3$ shapes, and $|C| \times |S|$ objects. x_i means the joint probability of the appearance, hence $0 \le x_i \le 1$ and $\Sigma x_i = 1$.

- In this study, we fix the number of the colors $|C| = 10$ and of the shapes $|S| = 20$, so there are 200 different objects.
- There are two agents: an infant and an adult. In each turn, the infant is taught 50 correspondences between labels and objects, chosen randomly, by the adult, and then tries to perform the task. There is no noise in the instruction: The adult does not teach wrong correspondences.

Table 3: An example of the matrix of joint occurrence rates between attributes (colors and shapes) and objects.

	○	△	□	●	▲	■
white	x_1	x_3	x_5	0	0	0
black	0	0	0	x_7	x_9	x_{11}
circle	x_2	0	0	x_8	0	0
triangle	0	x_4	0	0	x_{10}	0
square	0	0	x_6	0	0	x_{12}

4.2 Task 1: What is this Object?

The first task is to answer the question "What is this object?" after being taught some correspondences. The answering process "What" takes an object o as the argument and returns a label l that satisfies

$$\mathrm{What}(o) = \underset{l' \in L}{\mathrm{argmax}} B(o|l'), \tag{11}$$

where $l = \mathrm{What}(o), \mathrm{What} : O \to L$. If there are some labels equally probable to be answered, one is randomly chosen. Because there are two attributes, color and shape, the answer can be the object's color or shape. The answer is correct if it is one of the answers by the adult that is correct by definition,

$$\mathrm{What}(o) \in \mathrm{What_a}(o), \tag{12}$$

where "$\mathrm{What_a}(o)$" denotes the set of correct answers to the "What is o?" question by adults. "$\mathrm{What_a}$" operator has the type of $O \to \mathcal{P}(L)$. For example, $\mathrm{What_a}(\blacktriangle) = \{\mathrm{black}, \mathrm{triangle}\}$. For any o in O, $|\mathrm{What_a}(o)|$ is the number of attributes. In this study, the number of attributes is 2 (color and shape). The infant is asked for answers for all objects in O. The index shown in the Results section is the correct rate, the proportion of the correct answers for $|O|$ questions.

4.3 Task 2: Which Object has this Label?

The second task is to answer the question "Which is it?," with a label l presented. It is to answer an object o that satisfies

$$\mathrm{Which}(l) = \operatorname*{argmax}_{o \in O} B(l|o), \mathrm{Which} : L \to O \tag{13}$$

where $o = \mathrm{Which}(l)$. As in task 1, one object is randomly chosen if there are some objects equally likely. The answer $o = \mathrm{Which}(l)$ is correct if it satisfies

$$\mathrm{Which}(l) \in \mathrm{Which_a}(l), \tag{14}$$

where the type of $\mathrm{Which_a}$ is $L \to \mathcal{P}(O)$. For example, $\mathrm{Which_a}(\mathrm{circle}) = \{\circ, \bullet\}$ and $\mathrm{Which_a}(\mathrm{white}) = \{\circ, \triangle, \square\}$. $\forall l \in L, |\mathrm{Which_a}(l)| = n$ holds, where n denotes the number of the other attribute than of l. The adult asks the infant a question for all labels in L. In this task the correct rate index is the proportion of the correct answers for $|L|$ questions.

4.4 Task 3: Give me the Thingy!

The preceding two tasks just test the memory retention, so they may appear to be trivial, though we will see that the biases, when rigid as in RS, work as obstacles in the Results section. The third task tests the value of learning that is in its use. So we define another task that is to request an object that is not at the infant agent. This task is mathematically similar to the calculation of a fixed point. Let r be a probabilistic operator that randomly chooses one element from a set. Then the criterion of this task is written as:

$$r(\mathrm{Which_a}(\mathrm{What}(o))) = o, \mathrm{Which_a} \circ \mathrm{What} : O \to \mathcal{P}(O). \tag{15}$$

For example, an infant wants the white circle. However, because it is not present in front of her, she must tell one of its labels, white or circle. (Compound nouns are not allowed here.) If she presents the label white, the adult will bring her something white. It has the shape of circle, the infant finally gets what she wants.

This task is, however, too difficult. If the adult uniformly randomly chooses one white object from the white ones, the probability that the infant gets the circle is $1/|S|$. It is $1/3$ in the example in Table 3 and $1/20$ in the simulations.

So we take the following weaker version of the question:

$$\mathrm{What}(r(\mathrm{Which_a}(\mathrm{What}(o)))) = \mathrm{What}(o), \mathrm{What} \circ \mathrm{Which_a} \circ \mathrm{What} : O \to L \tag{16}$$

Then the task becomes easier. It is that the infant gets satisfied even when she has got the white triangle, if only the white triangle is "white" for her rather than "triangle", $B(\mathrm{white}|\triangle) > B(\mathrm{triangle}|\triangle)$. Mathematically, the object brought by the adult to meet the request of the infant, $\mathrm{Which_a}(\mathrm{What}(o))$ is correct if it is in the equivalence class defined by the map What.

Note that task 3 can not always be achieved even if task 1 and 2 could have been carried out at a satisfactory level. For example, if $\mathrm{Which}(l) = \mathrm{Which_a}(l)$ holds (though generally $\mathrm{Which_a}$ is a non-singleton set so a random selection mediates), the formula (15) becomes

$$\mathrm{Which}(\mathrm{What})(o) = o, \tag{17}$$

meaning that

if the infant answers l to "What is o," then she answers o to "Which is l".

Thus we see that some S bias helps to achieve this task.

5 Results

The results shown here are of the average of 1,000 simulations. The proportion of correct answers of NS, LS and RS in task 1, 2 and 3 are shown in Figure 3, 4 and 5, respectively.

NS can learn optimally in task 1 and 2. It fails only when the object of the label in question has been not yet learnt, so rapidly the correct rate converges to 100%. However, NS cannot communicate well in task 3. It can get what it wants everlastingly with a probability no more than 55%.

On the other hand, RS cannot execute task 1 and 2 that is just to answer the learnt contingency. Here we see that the complete S and MX biases of RS damage such simple procedures. It learns well but can communicate only with the poorly learnt vocabulary.

We see that LS satisfactorily execute all the tasks. In task 1 and 2, LS does even slightly better than NS. In task 3, LS is a little bit discreet than RS but steadily raises the correct rate according to the amount of learning. This is a good property since the learning and the tasks do not form a feedback loop: There is no reward nor punishment. LS does so without any instruction.

6 Discussion

Here we argue that LS is a model of development in the sense of Vygotsky [13].

6.1 Vygotskyan Development Where Bottom-up and Top-down Processes are Paradoxically Coordinated

Vygotsky defines development as a field where top-down and bottom-up processes intersect. The typical developmental processes are cited in Table 4. He focuses on the

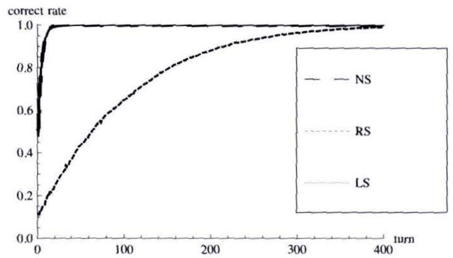

Fig. 3: The correct rate in task 1.

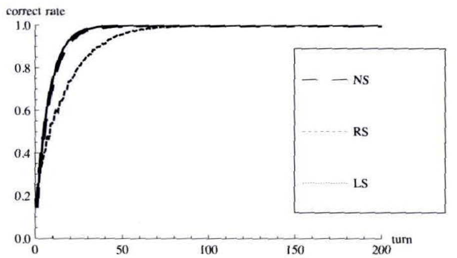

Fig. 4: The correct rate in task 2.

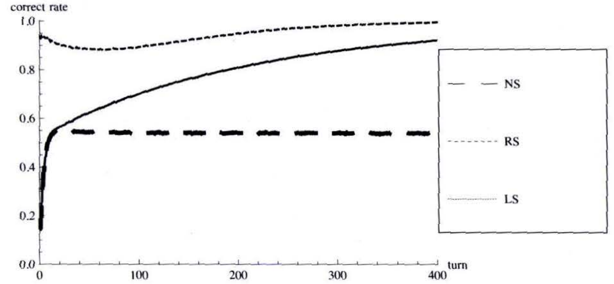

Fig. 5: The correct rate in task 3.

interaction of the two modes of learning. For example, we learn our mother language in an inductive way. We hear many sentences and we induce the grammar, the generative rules for utterance. The learning process goes from experiences to rules, in a bottom-up way. On the other hand, learning foreign languages progresses in a quite different, deductive, way. We first learn the alphabet, elementary vocabulary and the grammar. Then we begin to construct sentences. It is in a top-down way. The two ways, bottom-up and top-down, are not completely consistent. However, they become closely related and the close relationship drives the development. The agents in development, who are essentially internal observers, are required to make the two processes consistent, and sometimes the effort to make them consistent itself brings about inconsistency. It is an endless process.

Table 4: Examples of top-down and bottom-up developmental processes.

development in:	top-down	bottom-up
language acquisition	foreign language	mother language
concept	scientific/nonspontaneous	spontaneous
language	written	spoken

6.2 LS as the Model of Development

There are bottom-up and top-down forms in the probabilistic framework. NS estimates the causal intensity of (candidate) causes in an independent way. $NS(q|p) = a/(a+b)$ and $NS(q|\bar{p}) = c/(c+d)$ are just independent. The estimation by NS is absolute and bottom-up. On the other hand, RS satisfies a property that we call estimation relativity (ER). ER is derived from MX bias (eq. 1) and XM (eq. 3).

$$RS(q|p) = 1 - RS(q|\bar{p}). \qquad \text{(ER)} \qquad (18)$$

Because of this property RS's estimation is in a top-down way. In the calculation process, the invariant whole, 1.0, is given first and then it is divided into two parts, $RS(q|p)$ and $RS(q|\bar{p})$ that are the whole in sum total. The estimation by RS is relative and top-down. NS does not presuppose that there is such an invariant or conserved totality.

Our LS model can be considered as a realization of Vygotsky's notion of development. Here we mention only one simple reason. It is that LS is in between CP and RS. We consider that LS model exemplifies the importance of Vygotsky's theory of development or ZPD (zone of proximal development) theory. For example, in decision making, LS estimates the action value of an option, not solely absolutely (independently) nor relatively (dependently). Absolute and relative estimations are respectively bottom-up and top-down. The calculation of LS's value is a process where the heterogeneous ways of estimation, independent and dependent, meet.

7 Conclusion

In this study, we have shown the effectivity of LS model in word learning. It is that we have given LS back to its original realm, since LS is, from the first moment, an implementation of an adjustment mechanism of the intensity of S and MX biases.

Our LS model has variable S and MX biases, of which the intensity gets adjusted according to the situation or experience. The model is applied to word learning tasks by infant agents. By computer simulations, we see that (1) the biases exert adverse effect in the word learning tasks but (2) the agents should be somewhat biased to handle the communication task with the teacher agents, utilizing the acquired words. LS model accomplishes the both tasks very well, while most of agents with other models can satisfactorily execute only one of the two kinds of tasks.

Acknowledgements

The authors would like to thank Tokyo Denki University Science Promotion Fund for research grant 0509020001670, Japan Women's University for research project grant 41-2009, and Gakushuin University Computer Centre for research project grant 2009. Part of this work was carried out under the Cooperative Research Project Program of the Research Institute of Electrical Communication, Tohoku University.

References

[1] Dubois Daniel M. (2000). Review of Incursive, Hyperincursive and Anticipatory Systems - Foundation of Anticipation in Electromagnetism, *Computing Anticipatory Systems: CASYS'99 - Third International Conference. Edited by Daniel M. Dubois, Published by The American Institute of Physics, AIP Conference Proceedings* 517, 3–30.

[2] Gilovich Thomas, Griffin Dale W. and Kahneman Daniel (eds) (2002). *Heuristics and Biases: the Psychology of Intuitive Judgement*, Cambridge University Press.

[3] Hattori Masasi (2003). Adaptive Heuristics of Covariation Detection: A Model of Causal Induction. In: *Proceedings of the 4th International Conference on Cognitive Science and the 7th Australasian Society for Cognitive Science Joint Conference (ICCS/ASCS 2003)* , 1, 163–168.

[4] Hattori Masasi and Oaksford Mike (2007). Adaptive Non-Interventional Heuristics for Covariation Detection in Causal Induction: Model Comparison and Rational Analysis, *Cognitive Science* , 31, 765–814.

[5] Hattori Masasi (2008). The equiprobability hypothesis of reasoning and judgment: Symmetry in thinking and its adaptive implications, *Cognitive Studies*, 15(3), 408–427. (in Japanese)

[6] Markman Ellen M. and Wachtel Gwyn F. (1988). Children's Use of Mutual Exclusivity to Constrain the Meaning of Words. *Cognitive Psychology* , 20, 121–157.

[7] Rosen Robert (1985). *Anticipatory Systems* , Pergamon Press.

[8] Shinohara Shuji, Taguchi Ryo, Katsurada Kouichi, and Nitta Tsuneo (2007). A Model of Belief Formation Based on Causality and Application to N-armed Bandit Problem, *Transactions of the Japanese Society for Artificial Intelligence*, 22, 1, 58–68. (in Japanese)

[9] Sidman Murray and Tailby William (1982). Conditional Discrimination vs. Matching to Sample: an Expansion of the Testing Paradigm, *J. Exp. Anal. Behav.*, 37, 5–22.

[10] Spencer-Brown George (1969) *Laws of Form*, George Allen & Unwin, London, 1969.

[11] Takahashi Tatsuji, Masahiro Nakano and Shuji Shionohara (2010). Cognitive Symmetry: Illogical but Rational Biases, *Symmetry: Culture and Science* , 21, 1-3, 275–294.

[12] Takahashi Tatsuji, Oyo Kuratomo and Shuji Shionohara (to appear). A Loosely Symmetric Model of Cognition, *LNCS / LNAI* 5777, 5778.

[13] Vygotsky Lev S. (1986). *Thought and Language* , The MIT Press.

[14] Yamazaki Yumiko (2004). Logical and illogical behavior in animals, *Japanese Psychological Research* , 46, 3, 195–206.