ベイズ的枠組みにおける対称性推論のモデル化と認知バイアスの体系化に向けて

Towards a modeling of symmetry reasoning and a systematic understanding of cognitive biases in Bayesian frameworks

大用 庫智[†]高橋 達二
Kuratomo Oyo, Tatsuji Takahashi

東京電機大学
Tokyo Denki University
Kuratomo.oyo at gmail.com, tatsujit at mail.dendai.ac.jp

Abstract

We aim to clarify the basics of human cognition and focus on the loosely symmetry (LS) model, a kind of conditional probability flexibly that autonomously adjusts symmetry and mutual exclusivity. We showed that it is an empirical Bayes model which has a beta distribution both in prior and posteriori distribution. LS describes the data of causal induction with no hierarchization nor external estimation of distribution. In the decision-making context, two biases, the gambler's fallacy and the hot hand fallacy, could be expressed by the probability distributions that LS assumes. From the above, we suggested that the basis of human cognition may be characterized by prior distribution, and a possibility that the potential part of human cognition could be treated quantitatively in this way.

Keywords — causal induction; n-armed bandit problems; gambler's fallacy; hot hand fallacy; symmetry bias; mutual exclusivity bias

1. はじめに

近年、人間の認知、特に推論や判断に、規範的な合理性に従わない偏り（認知バイアス）が存在することが明らかとなり、多種多様な認知バイアスの分析に伴った人間認知の研究が進んでいる。しかしながら認知バイアスの相次ぐ発見と定式化に対し、その体系化は充分に進んでいるとは言えず、博物学的段階からの脱却をもたらす理論化が

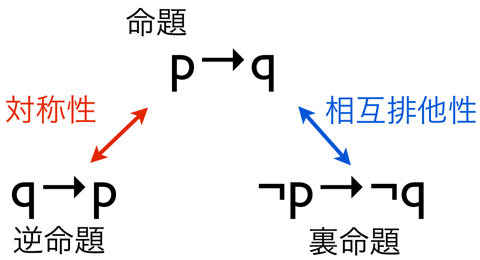

図 1：対称性と相互排他性

待たれている。そのため我々は、対称性推論、特に対称性と相互排他性と呼ばれている認知バイアスを人間の認知に根本的なものと仮定した研究を続けてきた（[11]-[13], [15]-[22]）。本研究では、これまでの研究のベイズ的意味合いを明らかにすることを試みる。

図 1 の様に対称性と相互排他性は、条件文からそれぞれその逆と裏を想起する認知の偏りである。対称性については、今井が発達心理学の分野で重要性を唱え[8]、また服部が思考心理学における広範な存在を示し、確率論的観点から等確率性仮説として定式化し、さらに情報理論的な観点からの正当化を行っている[7]。現在、対称性と相互排他性はほぼ人間に固有なものと考えられており、幼児の言語獲得には必要不可欠であり、言語や記号化、コミュニケーションなどを支えるものと議論されている。また、服部は対称性に着目して高次認知のモデルを提唱している。たとえば因果帰納とは事象間の共変動情報から因果関係を帰納的に形成する推論であるが、これに関して服部は DFH モデルを提唱し[5]、その有効性を証明している。

表 1: 2×2 の分割表の共変動情報

		結果	
		q	$\bar{q}$
原因	p	a	b
	$\bar{p}$	c	d

a：p と q が共に発生した回数
b：p と $\bar{q}$ が共に発生した回数
c：$\bar{p}$ と q が共に発生した回数
d：$\bar{p}$ と $\bar{q}$ が共に発生した回数

$$DFH = \sqrt{P(q|p)P(p|q)} = \frac{a}{\sqrt{(a+b)(a+c)}} \qquad \text{式 1}$$

$$LS(q|p) = \frac{a + \frac{bd}{b+d}}{a + \frac{bd}{b+d} + b + \frac{ac}{a+c}} \qquad \text{式 2}$$

$$\Phi = \frac{ad - bc}{\sqrt{(a+b)(c+d)(a+c)(b+d)}} \qquad \text{式 3}$$

DFH は式 3 の四分点相関係数（Φ）のdセルを無限大に発散させた形式であり、観測の対象となる事象の発生が比較的に稀であるという稀少性（a, b, c[1]が稀少）を前提としている。この稀少性は人間がデフォルトで仮定していると考えられており、人がdセルよりもa, b, cセルの情報をより有益と捉えるとされている。また、稀少性を仮定した時に、少なくともベイズ的な観点からは、dセルよりもa, b, cセルがより有益である[1]。

人間の推論は背景知識に手引きされるため、人間の認知の基礎を考えるためには、背景知識による不確実な統計情報の扱いに関しても議論しなければならない。このことに関し、Tenenbaum らは事前分布を階層的に扱うことで、その抽象的な知識を特徴づけられるとしている[3, 9]。そういったベイズ的分析とモデリングを行う場合には通常、超パラメータの階層化や推定が必要である。

そこで我々は、認知バイアスの理論化を通じた人間認知の基礎の解明を目指し、対称性と相互排他性を自律的に柔軟に調節するという観点から考案された緩い対称性（*LS*）モデル[18]に着目してきた(式 2)。このモデルは、人間の因果帰納の傾向を DFH より良く記述し、かつ意思決定などの多くの課題において人間と類似した高いパフォーマンスを見せる。幼児の言語獲得[11]や強化学習[16]等で優秀な成績と高い汎化能力を示し、いくつもの認知バイアスを再現している。ただし、*LS* の定義の必然性については、高橋らによってゲシュタルト心理学における図と地に例えた分析があるものの[21]、理論的分析は未だ不十分であった。

2. 緩い対称性モデルと経験ベイズ

そこで本研究では、事前分布のパラメータを経験的に決定する経験ベイズ法を利用して、LS に理論的な背景を与え、そのベイズ的枠組みを用いて

事後分布(ベータ分布) ∝ 事前分布(ベータ分布) × 二項尤度関数

$$g(\theta_i|\gamma_1 + a, \gamma_2 + n_i - a) = \frac{g(\theta_i|\gamma_1,\gamma_2)f(a|n_i,\theta_i)}{\int_0^1 g(\theta_i|\gamma_1,\gamma_2)f(a|n_i,\theta_i)d\theta}$$

事後分布から期待値を取る

$$LS(q|p) = \frac{a+\gamma_1}{a+\gamma_1+n_i-a+\gamma_2} = \frac{P(p,q)+P(p,\bar{q})P(\bar{p}|\bar{q})}{P(p)+P(p,\bar{q})P(\bar{p}|\bar{q})+P(p,q)P(\bar{p}|q)}$$

$n_i = a + b$
$N = a + b + c + d$
γ_1：パラメータ=$P(p,\bar{q})P(\bar{p}|\bar{q})/N = bd/(b+d)$
γ_2：パラメータ=$P(p,q)P(\bar{p}|q)/N = ac/(a+c)$

図 2：LS モデルとベイズの定理による理論的な背景

因果帰納の推論と意思決定のタスクを分析する。その分析に扱う *LS* (式 2) は、**表 1** の分割表上の共変動情報 a, b, c, d の関数として定義され、最も簡単な条件付き確率に対称性と相互排他性の調節項（図 2 の $\gamma_1 = \frac{P(p,\bar{q})P(\bar{p}|\bar{q})}{N} = bd/(b+d)$ と $\gamma_2 = \frac{P(p,q)P(\bar{p}|q)}{N} = ac/(a+c)$）を加えた形式である。$N$ は共変動情報の頻度の合計、 $a+b+c+d$ である。これらの LS の調節項は、バイアスを持たない式 **4** の条件付き確率と対称性と相互排他性を完全に満たす式 5 の RS モデルとの間で対称性と相互排他性を自律的に調節する[18][21]。

$$CP(q|p) = \frac{a}{a+b} \qquad \text{式 4}$$

$$RS(q|p) = \frac{a+d}{a+b+c+d} \qquad \text{式 5}$$

ここで扱う問題空間が二項分布であるため、[3] で扱われている事前 $(g(\theta_i|\gamma_1,\gamma_2))$と事後分布 $(g(\theta_i|\gamma_1 + a, \gamma_2 + \mathrm{n_i} - a))$がベータ分布になる二項–ベータモデル（図 2 のベイズの式）を LS の導出に用いる（ここでベータ分布が用いられるのはその利便性によるもので、ベイズ式の計算が容易でありかつ、様々な形状の分布を表現できる）。図 2 の$g(|)$はベータ分布の確率密度関数、$f(|)$は二項分布の確率密度関数であり、それぞれ分布を特徴づけるパラメータを持つ。上記の確率密度関数とベイズの定理の関係により、事前分布と事後分布の期待値[2]は、$\frac{\gamma_1}{\gamma_1+\gamma_2}$と$a+\gamma_1/(a+\gamma_1+\mathrm{n_1}-a+\gamma_2)$になる。図 2 の様に事前分布のパラメータに調節

[1] a, b, c セルは原因または結果の何れかが発生しているが、d セルは何も起きなかったという事を意味している。

[2] γ_1を a、γ_2を b とした時、ベータ分布の期待値は，$a/(a+b) = P(q|p)$となる。

項を代入して事後分布から期待値を取ると式 2 の LS を導かれることが大用の修士論文 [13]で明らかになった。

以上の様に LS は共変動情報から簡単に計算出来、背景に二項ベータモデルを持つ。そのため、人間が因果帰納や意思決定を行う際の背景知識（事前分布）と事後分布の関係をベータ分布で表現することが可能になった。以降の章では、LS モデルが因果帰納のデータに適合する事を示す。そして、LS の背景にあるベイズ的枠組み（確率分布間の関係）を用いて、なぜデータに適合するのかを示し、因果帰納を行う人間の傾向・態度から意思決定でよく観られる二つのバイアスを表現できる可能性を示す。

4. 因果帰納へのデータの適合

因果帰納の実験では、被験者は原因と結果の在不在の組合せを受動的に観測し、その後原因と結果の関係性の強さを尋ねられる。被験者の原因と結果の観測は、**表 1** の共変動情報 a, b, c, d の回数に従い、原因と結果の画像を PC のディスプレイを通して行われた。また、被験者は原因と結果の関係性の強さを 0 から 100 の範囲のスライダーで回答した。**表 1** の共変動情報の組合せを数種類観測した後、その組合せ毎の人間の評価値とモデルの評価値を用いて相関係数と誤差を計算する。

分析には LS と DFH, 式 3 の四分点相関係数(Φ)の三つのモデルを用いた。また、以下の三つの実験データを扱った。原因と結果を無相関に捉えるかを調査した実験[15]と[5]で行われた因果帰納のメタ分析、そして[14]の実験条件を拡張した複数要因[15]の結果を表 2 に示す。

表 2：因果帰納における分析結果

	LS	DFH	Φ
無相関	(.83, 8.1)	測定不可能	(.23, 19)
複数要因	(.99, 3.8)	(.98, 10.1)	(.92, 26)
メタ分析	(.96, 9.09)	(.95, 10.44)	(.86 11.5)

データの組合せの傾向： d > a > c > b

パラメータの重み： α> β > γ > δ

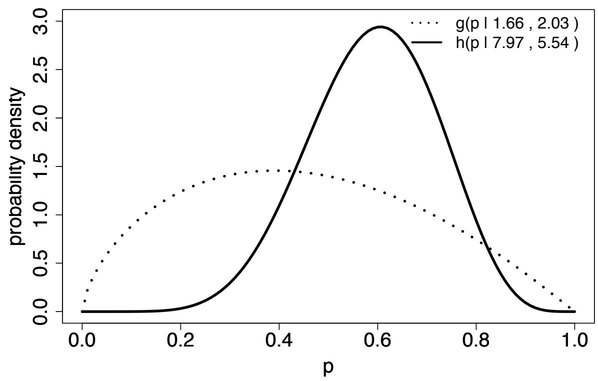

図 3：因果帰納実験における LS の事前分布と事後分布の関係。実線が事後分布。点線が事前分布。

表 2 から LS は最も相関が高く、誤差が低いため人間の因果帰納の傾向が LS に近いことが分かる。

なぜ LS が因果帰納のデータに適合するのかを分析するために、刺激の組合せ数が最も多いメタ分析のデータからデータの組合せの傾向（a = 6.31, b = 3.51, c = 4.56, d=6.85）を表 2 に示す。つまり、因果帰納の実験では、データの母集団が$P(\bar{p},\bar{q}) > P(p,q) > P(p,\bar{q}) > P(\bar{p},q)$という傾向を持つ事と原因と結果がどちらも起きなかったという d セルの観測が最も多いことが分かる。

上記を分析するための方法として、[10]の線形結合(EI)モデルを用いたパラメータ分析と[1, 2]のベイズ的枠組みからの分析の両方を用いる。式 6 の EI モデルは、式 5 にα、β、γ、δという四つの自由パラメータを持つ形式であり、各パラメータは[0, 1]の範囲の数値である。分析は各自由パラメータを 0.01 刻みで全ての組合せを調査した。これらのパラメータに $\alpha = 1,\ \beta = 1,\ \gamma = ac/(a+c), \delta = bd/(b+d)$と代入すると LS モデルと等しくなる。

$$EI(\alpha,\beta,\gamma,\delta) = \frac{\alpha a + \delta d}{\alpha a + \beta b + \gamma c + \delta d} \qquad 式\ 6$$

この EI モデルを用いて人間のデータと相関係数が最も高くなるパラメータ（α= *1.0*、β = *0.9*、γ= *0.5*、δ = *0.1*）をを表 2 に示す。データの母集団が$P(\bar{p},\bar{q}) > P(p,q) > P(p,\bar{q}) > P(\bar{p},q)$ である一方、人間は a の観測（p と q が共に発生）を

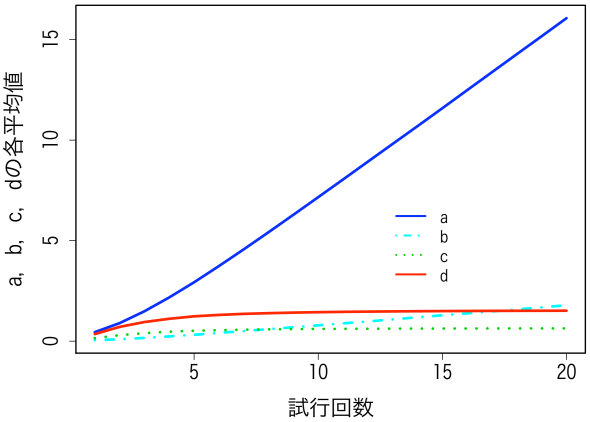

図 4：LS による二本腕バンディット問題（$P(q|p)=0.9, P(q|\bar{p})=0.3$）の実行における a, b, c, d の平均的時間発展。

重視している事が分かり、[10]の分析で使われたデータと同様な結果を示した。共変動情報の研究の成果として、人間は原因と結果の共通の不在(d)よりも共通の在 (a) を重視することが分かっている[1, 2]。Mckenzie はこれらの傾向を稀少性（P(p)と P(q)が低い）の元で尤度比を用いて分析し、少なくともベイズの観点から a を獲得することが有益である事を示している[1, 2]。以上の知見を元に LS の背景と因果帰納の関係を議論する。因果帰納のデータの母集団の傾向から LS の事前と事後分布の関係を図 3 に示す。図 3 から因果帰納のデータの母集団の傾向が $P(\bar{p},\bar{q}) > P(p,q) > P(p,\bar{q}) > P(\bar{p},q)$ であるのに対し、事前分布が確率が低い部分に集中しており、事後分布と差がある事が分かる。つまり、LS は事前の知識（背景知識）として、b セルが起こりやすいと考えており、その事前知識から影響され、因果関係の強さ（因果強度）を測るために使う事後分布は a セルが起こりやすいとしている。この様な分布間の関係から、LS は背景にある事前と事後分布を調節して a からの獲得する情報を重視しようとする戦略を持っている事が分かる。以上の考えは、人間の推論の多くは観測の対象となる事象は比較的に稀であるという稀少性と情報量の観点から正当化され、*LS* は人間同様、事前分布を用いることにより環境に対して適応的な戦略をとった結果、データに適合したと言える。

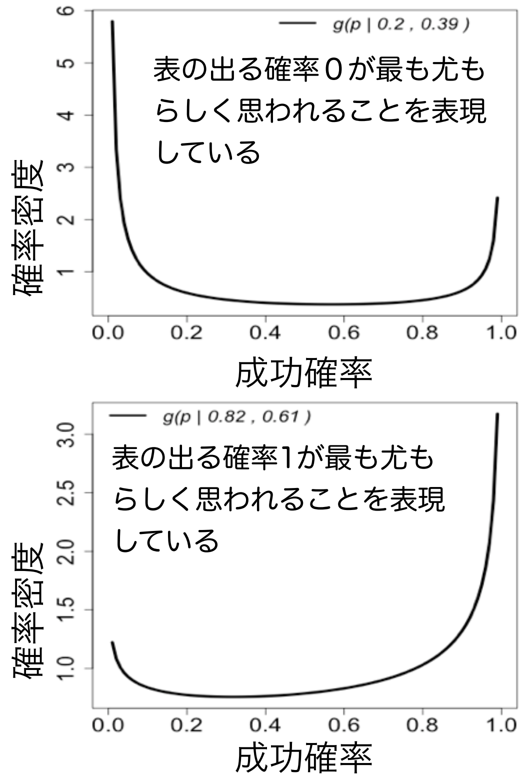

図 5：LS の事前分布の推移。4 回と 20 回試行後の共変動情報の平均は（a, b, c, d）=（2.18, 0.24, 0.47, 1.12）と（16.06, 1.78, 0.64, 1.51）である。

5. 認知バイアスの表現

意思決定のタスクとして、LS モデルが高いパフォーマンスを示している二本腕バンディット問題を扱い、確率分布（事前分布）で二つの認知バイアスを表現できることを示す。バンディット問題は、強化学習の最も基本的な問題であり、報酬確率が未知の二つのスロットマシンから得られる利益を最大化させる意思決定の問題である。ここで扱う二本腕バンディット問題の設定は、報酬確率が 0.9 と 0.3 であり、結果は一万回繰り返した平均である。以上の様な設定の場合、LS は数回の探索で報酬確率 0.9 のスロットを選択し続ける事が分かっている（つまり、a と b の値のみが加算され続け、LS の事後分布は最も大きい報酬確率に近い値を取る）。

そこで、ここでは、二本腕バンディット問題をコイントスに類比させ、LS とベイズ的な枠組みを用いて、（表が出やすく）偏ったコイントスの観測を仮定して、そこから事象（表）の発生に偏りがあるとその逆（裏）が出やすくなると期待してし

まうギャンブラーの誤謬から、その発生の偏りがさらに続くと今後もその事象（表）が発生すると考えるホットハンド[3]の発生への流れを表現する。

上記のバンディット問題の設定での LS の行動から求めた平均的な共変動情報 (a, b, c, d) を図 4 に示す。共変動情報 a (c)は 表が出る確率が高い（低い）コインを投げて表が出た時に 1 が追加され、 b (d) は表が出る確率が高い（低い）コインを投げて裏が出た時に 1 が追加される。図 4 から少ない試行回数から片方の腕に集中している事が分かる。図 4 の共変動情報を用いて計算した LS の事前分布を図 5 に示す。図 5 のベータ分布は、成功確率θ_1を[0, 1]の範囲で 0.01 刻みとして計算した。以上を LS の事前分布を通して観測すると、背景知識が観測前には U 字型になる。4 回試行の観測後では，事後分布の期待値(LS(q|p))は 0.79 という値を取りつつも左に歪み（図 5 上)、20 回試行の観測後では右に歪む（図 5 下、事後分布の期待値は 0.88)。つまり、観測前では表裏どちらがでても不思議ではないと考え、4 回試行後では裏がそろそろ出ると考えているが、20 試行後では今後は表が出続けるという背景知識の変化の流れを確率分布で表現している。

6. 議論

本研究では、対称性と相互排他性を自律的に調節する緩い対称性モデルにベイズ理論を背景にした導出を与えた。また、その背景から対称性と相互排他性を調節する調節項を用いて、獲得した頻度情報から自律的に確率分布を調節していることが明らかになった。LS は経験ベイズの枠組みを用いたモデルであるため、問題群の母集団を特定する様な形での分布の深い階層下やパラメータ推定を行っていない。それでも LS は因果帰納のデータに対して最も適合した。これは頻度情報から事前分布を用いて背景的に稀少性を仮定して、因果強度を算出しているためだと考えられる。また、意思決定の課題においては、ギャンブラーの誤謬とホットハンドという二つの認知バイアスを確率分布によって表現した。LS はバイアスを表現で出来るだけでなく、高い記述性能を持ちながら理論的な実験予測などが可能になると考えられる。

我々は、今後このように、*LS* モデルを対称性推論のトイモデルとして扱い、対称性推論とベイズ理論との関係をより明らかにし、また認知バイアスの由来として対称性や相互排他性を考えることが有用であることを示していきたいと思う。この主張に関しては今後多くの実験と検証を必要とするが、人間認知の基礎を事前分布で特徴付け、人間認知の潜在的な部分を定量的に扱うことで、より精細な分析と理論化、そして広範な分野への適用ができると思われる。

参考文献

[1]. Mckenzie, C. R. M. (2004), Framing effects in inference tasks and why they are normatively defensible. *Memory & Cognition*, 32 (6), 874-885.

[2]. Mckenzie, C. R. M., and. Laurie A. Mikkelsen. (2007). A Bayesian view of covariation assessment, *Cognitive Psychology,* 54, 33–61.

[3]. Griffiths, T. L., Kemp, C., and Tenenbaum, J. B, (2008), Bayesian models of cognition, In Ron Sun (ed.), *Cambridge Handbook of Computational Cognitive Modeling*, Cambridge University Press.

[4]. Gilovich, T., R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cognitive Psychology*, 17, 295-314.

[5]. Hattori, M., and Oaksford, M, (2007), Adaptive noninterventional heuristics for covariation detection in causal induction:Model comparison and rational analysis. *Cognitive Science*, 31(5), 765-814.

[6]. Gilovich, T., R. Vallone, and A. Tversky, "The hot hand in basketball: On the misperception of random sequences," *Cognitive Psychology*, 17, 295-314.

[7]. 服部雅史: 推論と判断の等確率性仮説, (2008), 思考の対称性とその適応的意味, *認知科学*,

15(3), 408–427.

[8]. 今井むつみ, 針生悦子. (2003), レキシコンの獲得における制約の役割とその性質, *人工知能学会誌*, 18, 1, 31–40.

[9]. J. B. Tenenbaum., C. Kemp., T. L. Griffiths., N. D. Goodman, (2011), How to Grow a Mind: Statistics, Structure, and Abstraction, *Science*, 331, 6022, 1279–1285.

[10]. Perales, J. C., and Shanks, D. R. (2007), Models of covariation-based causal judgment : A review and synthesis, *Psychonomic Bulletin & Review,* 14 (4), 577-596.

[11]. 神谷匠, 高橋達二. (2011). 緩い対称性による語彙学習バイアスの発現. *日本認知科学会第 28 回大会発表論文集*, 542-548.

[12]. 甲野祐, 高橋達二. (2010). 因果帰納の調和対称ヒューリスティクス. *日本認知科学会第 27 回大会発表論文集*, 43-46.

[13]. 甲野佑, 高橋達二. (2011). 緩い対称性モデルにおける不確定情報の扱い. *日本認知科学会第 28 回大会発表論文集*, 440-444.

[14]. Nichols, W., and Danks, D, (2007), Decision making using learned causal structures, *Proceedings of the 29th annual meeting of the cognitive science society*, 1343–1348.

[15]. 大用庫智, 高橋達二. (2010). 因果推論と意思決定を結ぶ緩い対称性モデル, *日本認知科学会第 27 回大会発表論文集*, 799-800.

[16]. 大用庫智, 甲野佑, 高橋達二. (2011). 非定常 N 本腕バンディット問題に対する人間の認知バイアスの適用, *2011 年度人工知能学会全国大会(第 25 回) 予稿集*, 1G1-2in.

[17]. 大用庫智. (2012). 不確定情報の評価と認知バイアスの確率的分布表現？：ベイズ主義の観点から, *2011 年度東京電機大学大学院理工学研究科 情報学専攻 修士論文.*

[18]. 篠原修二, 田口亮, 桂田浩一, 新田恒雄. (2007). 因果性に基づく信念形成モデルと *N* 本腕バンディット問題への適用, *人工知能学会論文誌*, 22, 1, 58–68.

[19]. 篠原修二, 田口 亮, 橋本敬, 桂田浩一, 新田恒雄. (2007). 語彙学習エージェントにおけるバイアスの自律調整について *人工知能学会論文誌*, 22(2), 103-114.

[20]. Takahashi, T., Shinohara, S., Oyo, K., and Nishikawa, A. (2011). Cognitive Symmetries as Bases for Anticipation: a Model of Vygotskyan Development Applied to Word Learning, *International Journal of Computing Anticipatory Systems*, 24, 95-106.

[21]. Takahashi, T., Nakano, M., and Shinohara, S., Cognitive Symmetry, (2011), Illogical but Rational Biases, *Symmetry: Culture and Science,* 21, 1-3, 275-294.

[22]. Takahashi, T., Oyo, K., & Shinohara, S, (2011), A Loosely Symmetric Model of Cognition. *Lecture Notes in Computer Science*, 5788, 234-241.