

眼球運動を用いた仮説生成における探索方略の検討

An Investigation of Search Strategies for Hypothesis Generation Using Eye Movement Data

松室 美紀[†], 三輪 和久[†]
Miki Matsumuro, Kazuhisa Miwa

[†] 名古屋大学大学院 情報科学研究科
Graduate school of Information Science, Nagoya University
muro@cog.human.nagoya-u.ac.jp

Abstract

The hypothesis generation process is divided into two components: 1) identifying factors or variables and 2) defining the relation between selecting factor(s) and results or outputs. We investigated how participants manage these two components in a rule discovery task using eye movement data. We found that the participants were categorized into two groups. The participants in one group supposed one particular factor as determining results and then consider the relations or regularities. The participants in the other group did not supposed one particular factor in the hypothesis generation process. Those in former group found the rule more frequently than those in later group.

Keywords — Rule discovery, Hypothesis generation, Eye tracking

1. はじめに

規則推論や(科学的)発見は日常において欠かせないプロセスの一つである。その過程は、仮説生成と仮説の検証からなる。先行研究において、その過程の様々な面が調べられてきた。例えば、仮説検証におけるバイアス[1]、仮説の生成と事例の生成(仮説の検証)の傾向[2]、実験の内容に関する研究[3]がある。しかし、発見、推論過程の大きな部分を占める、仮説生成の過程については、未だ研究が不足している。そこで、本研究では、仮説生成における方略に焦点を当てる。

仮説を生成するためには一般に、1) 結果とどの要因の間に、2) どのような関係が成り立っているか、という二つの点を決定する必要がある[4]。本研究では、前者を関連性の決定、後者を規則性の決定と定義する。関連性の検討、決定は仮説に使用する要因を検討し、ある要因が結果に関係すると仮定することである。要因を1つだけ選択し、主効果を仮定したり、複数の要因を選択し、要因間に交互作用があることを仮定したりする。規則性の検討、決定は要因と結果の間にどのような規則

性があるのかを検討、決定することである。規則性には、何らかの関係があるというような一般的なものから、 x が1増えると y が3増えるというような具体的なものまでが含まれている。例えば、ある状況において、“湿度を5%あげると、細菌aの増加率は2倍になる”という仮説を立てたとする。関連性の決定は、温度や湿度、明るさなどの様々な要因の中から細菌aの増加率という結果と湿度が関連していると仮定したことにあたる。規則性の決定は、仮定した要因(湿度)において、結果(細菌aの増加率)との間に5%の増加に対して2倍という規則性を仮定していることである。

規則生成過程に関する先行研究においては、規則性の検討、決定についての研究が主になされてきた。つまり、原因となる要因を明示した上で、その要因と結果の間に成り立つ関係を推論させていた。例えば、Haverty, Koedinger, Klahr and Alibali[5]は、 x と y という2つの値を示して、その間にどのような規則が成り立つかを推論させた。また、仮説生成支援の研究では、複数の要因を示して規則推論を行わせることが多いが、それらの研究では、仮説生成の過程については検討していない[4]。本研究では、複数の要因を示し、関連性の検討と規則性の検討を参加者がどのように行うのかを検討する。

1.1 眼球運動データの利用

これまでの仮説生成の研究は主に、プロトコルデータを用いて行われていた(e.g., [5, 6])。しかし、プロトコルデータにはいくつかの弱点がある。第一に、発言が思考に追いつかないという点である。そのため、全ての思考を発言することは困難である。また、思考の取得の時間分解能は粗くなってしまう。第二に、個人により発言量が異なる。人により、思考を言語的に表すことが得意であることとそうでないことがある。そして、悩んでいるとき、熟考しているときには思考を言語化できないことが多い。第三に、発言が規則の発見に影響

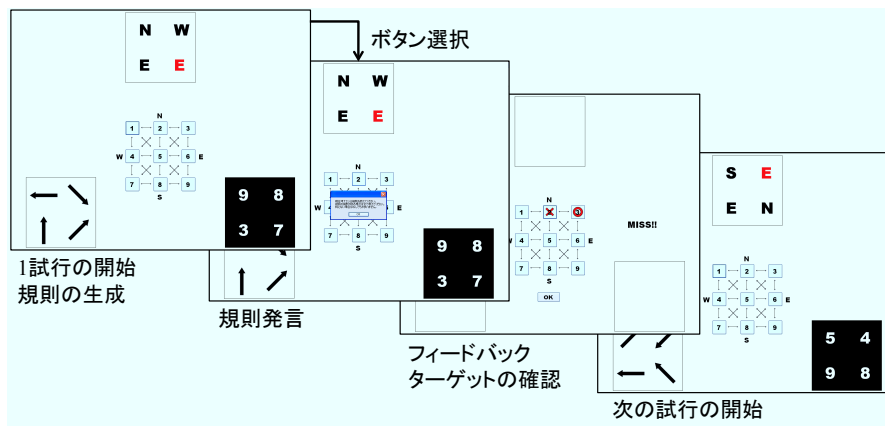


図1 本研究における一試行の流れの一例

表示されているオブジェクトは上のパネルが矢印、左下のパネルが方位、右下のパネルが数字である。中央の数字の書かれている9つの四角がボタンである。規則は課題を通して同じである。

を与えることがある。先行研究によると、説明活動を行うことにより、規則の発見が促進されるということが示されている [7]。逆の効果として、思考を発言することにより規則の発見が阻害されることも考えられる [8]。

そこで、本研究では、仮説生成過程の研究のために、新しく眼球運動測定を用いた実験パラダイムを提案する。眼球運動データは近年、様々な分野で利用されている。例えば、カテゴリー学習の研究においては、複数の要因にどのように注意を振り分けているかを検討するために、眼球運動データが使用された [9]。

1.2 課題

本研究では仮説生成に3つの要因が使用可能な課題を用いる。図1に課題の流れの一例を示す。課題画面は3つのパネルと9つのボタンからなり、各パネルには異なるオブジェクトが表示される。表示されるオブジェクトの種類は矢印、方位、数字である。9つのボタンのうち1つがある規則に基づいて、ターゲットに設定される。規則には3つのパネルのいずれか1つに表示されるオブジェクトを用いる。要因間の交互作用はないことが事前に教示される。参加者はできるだけ早く規則を見つけることを求められる。なお、規則発見の難易度を調整するため、オブジェクトの基本的な使用方法が教示された。

規則の生成、検証は以下の流れで行われる。参加者は3つのパネルに表示されたオブジェクトを基に、規則を考える。任意のタイミングで、9つのボタンのうち好きなボタンを押し、規則を検証することができる。ボタンを押すと、フィードバックに先立ち、参加者はどのような規則を保持しているかを発言するよう促される。規則の発言が終わる

と、ターゲットボタンの位置が表示される。ターゲットボタンの位置を確認後、参加者は再び、規則の検討を始める。このときパネルに表示されるオブジェクトは、前回の検討時とは数や方向、方位が変更されている (図1参照)。ただし、使用される規則は課題を通して同じ物である。参加者は7分の間、規則の生成、検証を何度でも行うことができる。規則の生成開始から、フィードバックまでを一試行とする。3試行連続で正しい規則に言及した場合、規則を発見したとし、課題を終了する。

本課題では、各パネルへの注視が各要因の検討と対応している。注視が異なるパネルへ移動したとき、異なる要因の関連性を検討しようとしているといえる。つまり、注視点の拡散から、一回の検証において、どのように関連性を仮定しているかを知ることができる。一回の仮説生成、検証 (1試行) において1つの要因の関連性のみ仮定し、規則性を検討している参加者は、関連性を仮定した一つの要因 (パネル) しか注視しないため、注視点は一つのパネルに集中する。1試行において複数の要因の関連性を仮定、検討する場合、関連性を仮定、検討した全ての要因 (パネル) を観察するため、注視点は複数のパネルに拡散する。本研究では、前者を注意集中探索、後者を注意拡散探索と定義する。

上記の課題を用いて、関連性と規則性の検討がどのように行われているかを検討する。以下の可能性が考えられる。第一に、人間は普遍的に同じ方略を使用する可能性である。この場合、全ての参加者の注視点の拡散は、集中か拡散のどちらかに収束する。対して、人により異なる方略を使用する可能性も考えられる。この場合、参加者の注視点の拡散は、その方略により異なる。また、人により異なる方略がとられた場合、それぞれの規

則発見率より、どの方略が規則の発見に有効か検討する。

2. 実験

2.1 方法

2.1.1 参加者

学部生22名が実験に参加した。

2.1.2 装置

課題は17インチ、1280×1024の解像度のモニタに表示された。参加者とモニタの距離は約60cmに設定された。各パネルの大きさは7.9°×7.9°、各パネルの中心はそれぞれ21.7°離れていた。眼球運動はTobii T60 eye trackerを用いて取得された。参加者は頭を自然に動かすことができる。

2.1.3 手続き

全参加者は課題を全部で5回行った。課題を行う前に、操作、および規則についての説明を受けた。その後、本番とは異なる規則で十分に練習を行った。本試行のためにeasy ruleとdifficult ruleが各要因に1つずつ、合計6つ用意された。easy ruleは練習に用いた規則とほぼ同じ簡単な規則、difficult ruleは練習に用いたルールとは少し異なる比較的难度の規則である。

はじめにeasy phaseとして3回の課題を行った。このフェーズでは各要因を用いたeasy ruleが1回ずつ用いられた。規則の順番、規則成立パネルの位置は参加者ごとに無作為に調整された。規則が成立するパネルの位置は3回を通して異なるようにした。その後、休憩をはさみ、difficult phaseが行われた。このフェーズでは各要因を用いたdifficult ruleの中からランダムに2つの規則が選ばれた。規則の順番、規則成立パネルの位置は無作為に調整したが、規則が成立するパネルの位置は互いに異なるように設定した。5回の課題を通して、各課題を行う前に、眼球運動測定のためのキャリブレーションを行った。また、規則の発見、未発見に関わらず、各課題終了後には各課題に使用されていた規則を教えた。

2.2 結果

2.2.1 使用方略の検討

始めに、各試行における参加者の注視点の拡散を算出する。用いたデータは、各参加者の各課題

における、予測失敗後の眼球運動である。つまり、予測失敗試行の次の試行の開始から規則の報告までの眼球運動である。予測成功後は、その時点で生成されている規則を保持することから、規則の生成を行わない。そのため、当該の規則が含まれる仮説セットに対応する1つのパネルに注意が集中するため、データとして用いなかった。各試行における各パネルへの注視時間の割合 (p_i) を求める。式(1)を用いて各課題における誤答タームの、adjusted-H (調整済みエントロピー) を産出した。

$$\text{adjusted-H} = -\frac{\sum p_i \cdot \log p_i}{\log 3} \quad (1)$$

このadjusted-Hの値が大きいほどその試行において複数のパネルへの注視があったことを示す。逆に、adjusted-Hの値が小さいほど、1つのパネルに注視点が集中していたこととなる。adjusted-Hの値は3つのパネルに均等に注視点が拡散されているときに最も大きい値である1をとり、いずれか1つのパネルに注視点が集中しているときに最も小さい値0をとる。

はじめに、課題間の使用方略に一貫性が見られるかどうかの分析を行う。この分析のために、課題ごとにadjusted-Hの平均を産出した。その相関を表1に示す。各課題間のadjusted-Hの平均の相関係数はいずれも0.5以上と高い相関を示している。このことから参加者は全ての課題を通して、一貫した方略を用いていることがわかる。特に、easy phaseの3問目以降は相関係数が0.8以上と非常に高い相関を示している。参加者の使用方略は一貫しているため、参加者ごとにどのような方略を用いているかを分類する。

参加者ごとにdifficult phaseの2回のadjusted-Hの平均値を算出する。全体の中央値は0.216となった。最小値は0.01、最大値は0.84、標準偏差は0.26であった。値の開きが大きく、参加者が複数のグループに分類できることが予測できる。ward法を用いたクラスター分析の結果、2つのグループに分類された(図2)。図2における上部10名は平均adjusted-Hが0.1以下であり、一回の検証で一つの要因に関連性を仮定したと考えられる。このグループを注意

表1 平均adjusted-Hの課題間の相関

	easy 1	2	3	difficult 1	2
1	1.000				
2	.729***	1.000			
3	.838***	.661**	1.000		
1	.752***	.620**	.889***	1.000	
2	.636**	.668**	.809***	.896***	1.000

** : $p < .01$, *** : $p < .001$

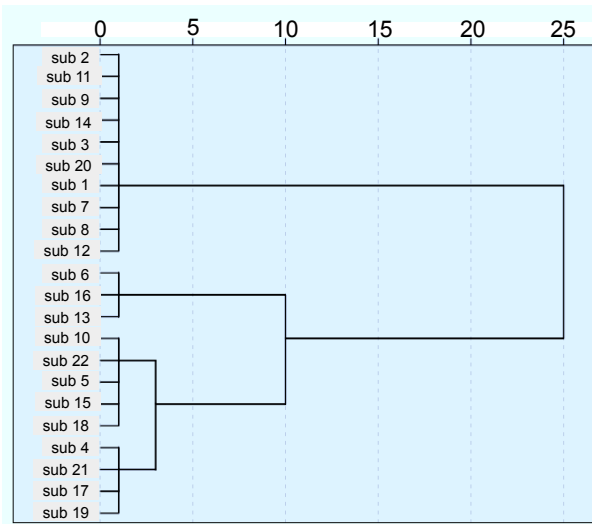


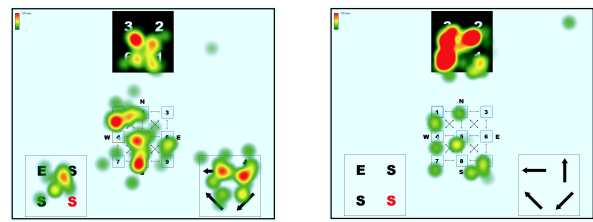
図2 クラスター分析の結果

集中グループと呼ぶ。もう一方のグループを注意拡散グループと呼ぶ。彼らは一回の検証において、複数の要因の関連性を検討したと考えられる。各グループのadjusted-Hの平均値を比較すると、有意な差が見られる ($t(11) = 6.730, p < .001$)。この結果から、仮説生成中の関連性と規則性の検討の関係は人により異なることが示された。

2つのグループの代表的な参加者の、ある1試行における注視点の拡散をヒートマップで表したものが、図3である。グループにより拡散が異なっていることが観察できる。また、課題中の時系列でadjusted-Hの変遷を表したものが図4である。各グループの参加者はそれぞれ2つの変遷パターンに分類される。注意集中グループの参加者はどちらのパターンも、値は基本的に0となっている。ほぼ全ての試行において、いずれか1つの要因を注視していることがわかる。対して、注意拡散グループの参加者は、複数の要因へ注視点が増加し、または、激しく上下している。2つのグループは根本的に異なる方略をとっていることが示された。

2.2.2 規則発見率

続いて、各グループにおける規則の発見率を比較する(表2)。両フェーズをあわせた全体的な規則発見率は、注意拡散グループ(43%)よりも注意集中グループ(68%)のほうが有意に高い($\chi^2(1) = 6.693, p = .010$)。さらに、フェーズごとに分けて細かく検討する。easy phaseにおいては注意拡散グループ(50%)よりも注意集中グループ(77%)のほうが有意に規則発見率が高かった($\chi^2(1) = 4.945, p = .026$)。しかし、difficult phaseにおいては2つのグループの規則発見率に有意な差は見られなかった($\chi^2(1) = 2.087, p = .149$)。そこで、difficult phase



(a) 注意拡散グループ (b) 注意集中グループ

図3 各グループの代表的な参加者のヒートマップ (a) 注意拡散グループの参加者のヒートマップ。複数のパネルに注視点が増加している。(b) 注意集中グループの参加者のヒートマップ。1つのパネルに注視点が増加している。

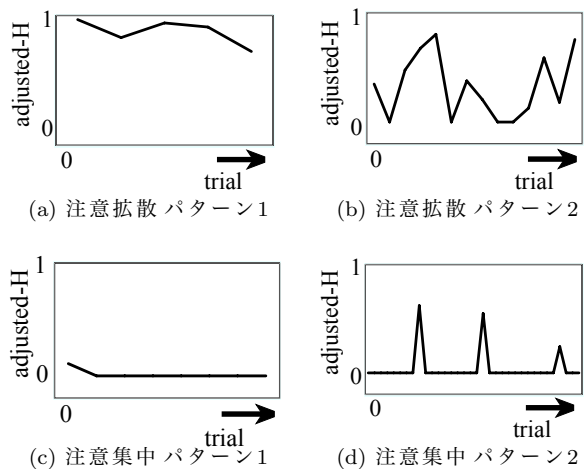


図4 adjusted-Hの課題中の変遷パターン

注意拡散グループにおける代表的な変遷2例を上段に、注意集中グループにおける代表的な変遷2例を下段に示す。各パネルは典型的な参加者の実際のパターンを示している。x軸は試行をy軸はadjusted-Hを表している。

の各課題を規則ごとに分けて分析を行う。すると、方位規則においては、注意拡散グループ(1/9)よりも注意集中グループ(4/6)において規則発見率が高かったが($p = .047$, Fischer's exact test)、矢印と数字規則においては有意な差は見られなかった($ps > .50$)。

矢印規則においては、注意拡散グループの規則発見者が8名中1名、注意集中グループの規則発見者が6名中1名と全体的に少なかったことから、床効果により有意差が見られなかったと考えられる。また、数字規則においては注意拡散グループの規則発見者が7名中6名、注意集中グループの規則発見者が8名中6名と全体的に多く、天井効果により有意差が見られなかったと考えられる。difficult phaseにおいては、上記の床効果、天井効果により

表2 各グループの各phaseにおける規則発見率

	注意拡散				注意集中			
	矢印	方位	数字	全体	矢印	方位	数字	全体
easy	0.17 (2)	0.42 (5)	0.92 (11)	0.50 (18)	0.70 (7)	0.90 (9)	0.70 (7)	0.77 (23)
difficult	0.13 (1)	0.11 (1)	0.86 (6)	0.33 (8)	0.17 (1)	0.67 (4)	0.75 (6)	0.55 (11)
全体	0.15 (3)	0.29 (6)	0.89(17)	0.43 (26)	0.50 (8)	0.81 (13)	0.72 (13)	0.68 (34)

()内は規則発見者数.

グループ間の差が観測されなかったといえる. 全課題, および, easy phaseにおける規則発見率, さらに, difficult phaseにおける適度な難易度であった方位規則における規則発見率から, 総合的に考えると, 1つの要因を仮定して, 規則性を検討する注意集中グループの参加者の方がより規則を発見したことが支持される.

2.2.3 生成規則と仮説生成方略

最後に, 参加者の関連性と規則性の検討方法と生成された規則の関係を検討する. 各試行において生成された規則は, 使用された要因と生成規則数から以下の3種類に分けることができる. 第一に, 一試行において, 1つの特定の要因に関連性を仮定し, 1種類の規則性(1つの規則)を生成した場合である. 例えば, 方位に関する規則を1つだけ生成した場合などである. 以下, 要因内単一仮説と呼ぶ. 第二に, 一試行において, 1つの特定の要因に関連性を仮定し, 複数の規則性(複数の規則)を生成した場合である. 例えば, 方位に関する規則を複数生成した場合がこれにあたる. 以下, 要因内複数仮説と呼ぶ. 第三に, 一試行において, 複数の要因に関連性を仮定し, 複数の規則性(複数の規則)を生成した場合である. 例えば, 方位に関する規則と, 数字に関する規則の両方を生成した場合がこれにあたる. 以下, 要因間複数仮説と呼ぶ.

参加者ごとに以下のように分類する. 課題を通して, 要因内単一仮説のみを生成し続けた参加者は要因内単一仮説の使用者に分類される. 要因内, 要因間複数仮説を一度でも生成した参加者はそれぞれ要因内複数仮説の使用者, 要因間複数仮説の使用者に分類される. 要因内, 要因間複数仮説の

両方を生成したことがある参加者は要因間複数仮説の使用者に分類される.

表3にグループごとの各規則タイプの使用者の人数を示す. 注意集中グループでは1試行において, 1つの要因にのみ注目するため, 規範的に, 要因間複数規則は生成され得ない. 単一規則の使用者に分類された参加者のほうが多い傾向が見られた($p = .055$). 一方, 注意拡散グループにおいて, 要因内複数規則の使用者に分類された参加者はいなかった. 参加者は, 要因内単一規則と要因間複数規則の使用者に偏りなく分かれた($p = .387$).

この結果は注意拡散グループにおいて異なる関連性の仮定法が用いられていた可能性を示す. ある参加者は複数の要因の関連性を検討し, その中から1つの要因に関連性を検討し, その要因においてのみ, 規則性の検討を行った. つまり, 最初に3つの要因全てを比較し, 方位に関連性を仮定してから, 規則性を検討したような場合である. 別の参加者は, 要因の関連性を仮定して, 規則性を検討するという流れを, 複数の要因において行ったと考えられる. 例えば, 方位に関連性を仮定し, 規則を生成. その後, 数字に関連性を仮定し, 規則を生成するという行動がこれにあたる. 結果的に, 前者の方法をとった参加者は要因内規則を, 後者の方法をとった参加者は要因間規則を生成した.

3. 考察

実験の結果, 参加者により要因の関連性の検討と規則性の検討の行い方が異なることが示された. これは, これまで規則性の検討しか行っていなかった, 規則発見の研究に, 要因の関連性の検討過程を含む必要があることを示唆している. 特に, 本課題においては, 関連性と規則性の検討の行い方により, 規則発見のパフォーマンスに影響が見られた. これは, 本実験で使用した課題において, 複数の要因の関連性を検討する利益がないためだと考えられる. 逆に, 複数の要因の関連性を検討することにより, 時間がかかり, 記憶容量の限界のため, 検討した情報が失われてしまう. そのため, 注意拡散グループにおける, 規則発見

表3 各規則タイプの使用者

	要因内 単一規則	要因内 複数規則	要因間 複数規則
注意集中	8	2	0
注意拡散	7	0	5

率が低かった可能性がある。ただし、この結果については異なる状況での検討が必要である。

本実験では参加者数が少なかったため、特に、注意拡散グループにおいて詳細な検討をすることができなかった。注意集中グループにおいては十分に小さい標準偏差が観察されたことから、全員がほぼ同様の行動をとっていると言える。しかし、標準偏差の大きさからみても、注意拡散グループには2種類のタイプの参加者が含まれていたと思われる。さらに、課題中の注視点の拡散の変異パターンや生成規則の内容も2つのパターンを示している。今後、参加者数を増やし、眼球運動の時系列的変化も含めた詳細な検討が必要とされる。

本実験の結果はdual space search theoryの観点からも検討できる。Klahr and Dunbar(1988)[10]によると、仮説空間は仮説のタイプごとにグループとなっており、異なるタイプの仮説への移行はより困難だとされている。大して、本実験においては、要因ごとに仮説がグループ化されていると考えられる。よって、注意集中方略は一度に仮説空間の狭い範囲を探索していると言える。一方、注意拡散方略は一度に仮説空間の広い範囲を探索していると言える。Klahrらの研究においては、一回の仮説生成でじっくり仮説を検討した参加者(theorist)のほうが、より早く、効率的に規則を発見していた。本実験では、一回の仮説生成で時間をかけて広い範囲を探索した、注意拡散グループの参加者のほうが、規則発見率が低かった。この結果は、仮説生成の際は、時間をかけて検討するときは、広い範囲を探索するよりも、ある程度の範囲に絞って探索する必要がある可能性を示している。ただし、本実験では全ての要因を事前に提示している点で先行研究と異なる。

実験課題の問題点として、フィードバックの回数が統制されていないことがある。2つのグループでボタンの押下数に有意な違いが見られたことから、規則発見率の違いはフィードバック数の違いによる影響である可能性がある。今後は、課題をより構造化し、条件の統制を行い、要因の関連性の検討と規則性の検討の関係性をさらに調べていく必要がある。

参考文献

- [1] Mynatt, C. R., Doherty, M. E., & Tweney, R. D., (1977) "Confirmation bias in a simulated research environment: an experimental study of scientific inference", *The Quarterly Journal of Experimental Psychology*, Vol. 29, pp. 85-95.
- [2] Klahr, D., (2000) *Exploring Science*, the MIT Press.
- [3] Tschirgri, J. E., (1980) "Sensible reasoning: a hypothesis about hypotheses", *Child Development*, Vol. 51, pp. 1-10.
- [4] Van Joolingen, W. R. & De Jong, T., (1991) "Supporting hypothesis generation by learners exploring an interactive computer simulation", *Instructional Science*, Vol. 20, pp. 389-404.
- [5] Haverty, L. A., Koedinger, K. R., Klahr, D., & Alibali, M. W., (2000) "Solving inductive reasoning problems in mathematics: not-so-trivial pursuit", *Cognitive Science*, Vol. 24, No. 2, pp. 249-298.
- [6] Qin, Y. & Simon, H. A., (1990) "Laboratory replication of scientific discovery processes", *Cognitive Science*, Vol. 14, pp. 281-312.
- [7] Williams, J. J. & Lombrozo, T., (2010) "The Role of Explanation in Discovery and Generalization: Evidence From Category Learning", *Cognitive Science*, Vol. 34, pp. 776-806.
- [8] Schooler, J. W., Ohlsson, S., & Brooks, K., (1993) "Thoughts beyond words: when language overshadows insight", *Journal of Experimental Psychology: General*, Vol. 122, No. 2, pp. 166-183.
- [9] Rehder, B. & Hoffman, A. B., (2005) "Eyetracking and selective attention in category learning", *Cognitive Psychology*, Vol. 51, pp. 1-41.
- [10] Klahr, D. & Dunbar, K., (1988) "Dual space search during scientific reasoning", *Cognitive Science*, Vol. 12, pp. 1-48.