

# 自律エージェントの利用する情報の顕著性が行動ルールの言語的・ 非言語的推定に与える影響

## The salience of information used by an autonomous agent on the estimation of the agent's action rules

二宮 由樹<sup>†</sup>, 下條 朝也<sup>‡</sup>, 寺井 仁<sup>§</sup>, 松林 翔太<sup>†</sup>, 三輪 和久<sup>†</sup>

Yuki Ninomiya, Asaya Shimojo, Hitoshi Terai, Shota Matsubayashi, Kazuhisa Miwa

<sup>†</sup>名古屋大学, <sup>‡</sup>コニカミノルタ株式会社, <sup>§</sup>近畿大学  
Nagoya University, KONICA MINOLTA, Kindai University  
[yninomiya@outlook.jp](mailto:yninomiya@outlook.jp)

### 概要

自律エージェントの行動を理解するには、何が入力情報かと、そこから生じる出力が何かを理解する必要がある。本稿は、この入力情報の顕著性が、エージェントに対する行動ルールの推定に与える影響を検討した。結果は、顕著性の高い情報と低い情報を入力情報として利用するエージェントの行動を言語的に推定する場合に、低い情報が利用されにくいことが示された。このことは、入力情報の顕著性によって行動ルールの推定における利用されやすさが異なることを示すとともに、行動ルールの推定における言語隠蔽効果の証拠を示すものである。

キーワード：ルール推定, 言語隠蔽効果, 自律エージェント

### 1. 問題と目的

近年、自律エージェントは目まぐるしく進歩をしている。これらの自律エージェントの行動を理解するためには、どのような情報が入力情報として利用されているのかを正しく理解したうえで、それがどのような出力につながるのかを理解する必要がある。

本研究はこの入力情報の顕著性(salience)の違いが、自律エージェントに対する行動ルールの推定に与える影響に着目する。我々は目の前に与えられた情報を等しく認識するのではなく、図と地を区別し、図の情報を優先的に認識している[1]。図と地の概念は知覚に関連するものだが推論や問題解決といった高次認知領域においても議論されることがある[2, 3]。本論では顕著性とはその情報の意識の向けやすさであると定義する。つまり顕著性が高いということは、その情報が図として意識を向けやすいということである。

ある。

一方で、様々な推論や問題解決課題において思考を言語化することでパフォーマンスが損なわれることが知られている[4]。このような現象を言語隠蔽効果という。言語隠蔽効果は、不適切な情報の顕著性の高い場合に、言語化によって、その情報に対してより意識を向けてしまうため生じると説明される[5]。このことから、情報の顕著性の高い情報と低い情報を入力情報とするエージェントの行動を推定する場合、言語化をすることで、顕著性の高い情報に注意を向けてしまい、低い情報が見逃される可能性がある。

本研究では、この言語隠蔽効果に着目し、言語化を伴う推定と非言語的な推定の2つの観点からエージェントに対する行動ルールの推定を行わせる。そして、入力情報の顕著性と言語化の関係を調べる。

### 2. 方法

**参加者** 参加者は108名(女性19名, 男性91名)で年齢は平均20.23( $SD=0.88$ )歳であった。実験はクラウドソーシングを利用してオンラインで実施した。

**手続き** 課題はjspsych[6]で作成したものを使用した。実験は、観察フェイズ、推定フェイズの2つのフェイズから構成されていた。

観察フェイズでは、参加者は迷路の中を移動するエージェントを観察した(図1a)。エージェントは迷路を移動中、周囲の状況に合わせて速度と色を4段階で変化させた(図2)。この変化は周囲の状況によって決定され、敵の数(0, 1, 2, 3以上)と逃げ道の数(4, 3, 2, 1)を入力情報とし出力(速度と色)が決定された。敵の数は画像の中にオブジェクトとして配置されているが、道の数には背景情報に着目する必要がある。そ

のため、敵の数のほうが逃げ道の数に比べて情報の顕著性が高いと考えられる。

観察前に参加者は、エージェントが周囲の状況に対して恐怖を感じ、色や速度を変化させると伝えられた。そして観察後にエージェントがどのような状況で恐怖を感じていたか答えてもらうため、状況に合わせて色や速度を変化させるエージェントを注意深く観察するようにと教示を受けた。そして、16本の動画を観察したのちに、推定フェイズを行った。

推定フェイズでは、非言語的推定課題と言語的推定課題を実施した。これらの課題順序は、言語隠蔽効果による、推定パフォーマンスの低下を考慮し、非言語的推定課題から実施した。

非言語的推定課題では、参加者は実験動画のスクリーンショット(図1b)を観察し、「この状況でエージェントはどの程度恐怖を感じていると思いますか」という問いに対し、シークバーを利用して恐怖度を0~100で回答した。つづく、言語的推定課題では、エージェントがどのような状況で恐怖を感じていたかについて「〇〇ならば恐怖を感じる」というIf文の穴を埋める形式で回答した。回答欄は制限なく用意されており、参加者は思いつく限りルールを記述することが出来た。

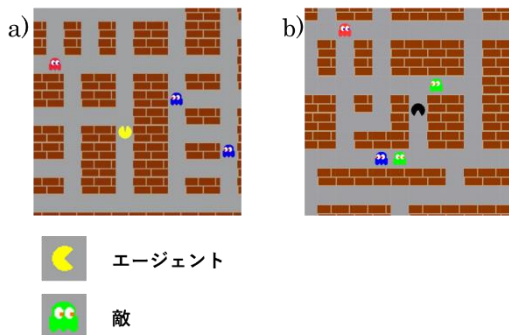


図1 観察フェイズのスクリーンショット(a)と非言語的推定課題における評価画像(b).



図2 周囲の状況に応じた出力(色)の変化

**刺激** 観察フェイズで使用する動画セットと非言語的推定課題で使用する評価用画像セットを作成した。動画セットを作成するにあたって、敵の数と逃げ道の数の全組み合わせ経験できるように、敵の数(4)×逃げ道の数(4)で計16パターンの状況を最低でもそ

れぞれ4回ずつ含むように動画を作成した。評価刺激セットも同様に、敵の数(4)×逃げ道の数(4)で計16パターンをそれぞれ4回分、すなわち計64パターンの評価刺激を作成した。

**実験計画** 入力情報の顕著性が行動ルールの推定に与える影響を検討するために、観察対象のエージェントが、顕著性の高い情報である敵の数によって出力(速さや色)を変える敵条件と、顕著性の低い情報である逃げ道の数によって出力を変える道条件を用意した。また、両方の情報を利用する条件として、出力を決定する際の重みが敵の数と未知の数の間で同じになるように設定した敵道条件を用意した。最後に、統制条件として、入力情報に対してランダムに出力を返す条件を用意した。

**予測** 言語隠蔽効果が意識にのぼりやすい情報に注意を向けることによって生じるという先行研究の知見[5]に基づき以下の予測が立つ。顕著性の高い情報を使用した行動ルールの推定は、非言語的推定課題においても、言語的推定課題においても正確に行えるだろう。一方で、顕著性の低い情報を使用した行動ルールの推定は、非言語的推定課題においては適切に行われるが、言語的推定課題では行われまいだろう。

### 3. 結果と考察

それぞれの推定課題において、参加者の回答から行動ルールの推定に、それぞれの入力情報をどの程度利用しているかの指標を作成し、分析した。

はじめに、非言語的推定課題では、評価をした状況における入力情報(敵の数、逃げ道の数)を説明変数、参加者の恐怖度評価を被説明変数とする以下のような重回帰式を参加者ごとに作成し、敵の数や逃げ道の数に対する偏回帰係数の値を最尤推定法により推定した。

$$\text{恐怖度} = \beta_{\text{enemy}} * \text{敵の数} + \beta_{\text{escape route}} * \text{逃げ道の数} + e$$

この時、 $\beta$ を行動ルールの推定時に、それぞれの情報をどの程度使用していたかについての重みと解釈した。各条件において、エージェントの行動ルールに使用される入力情報を正しく使用しているか検討するために各条件と統制条件の偏回帰係数を比較した。分析の結果、非言語的推定課題において、どの条件も、実際にエージェントが使用している入力情報に対する $\beta$ が統制条件よりも高かった(図3)。このことは、非言語的推定課題においては条件によらず、顕著性の高い情報も低い情報も使用して行動ルールを推定できていたことを示唆する。

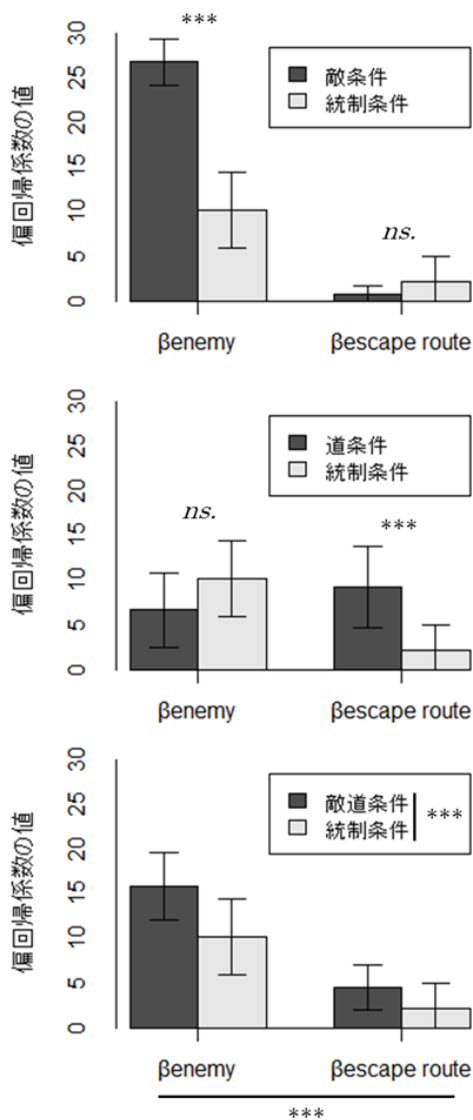


図3 重回帰式における偏回帰係数の統制条件との比較。上図，中図は交互作用があり(上: $F(1, 53) = 68.11, p < .001$ ，中: $F(1, 53) = 8.92, p < .01$ )，単純主効果検定を実施した結果を記載した(\*\*\*: $p < .001$ , ns.: $p > .10$ )。下図は，交互作用がなく( $F(1, 52) = 1.42, p > .10$ )，条件間・ $\beta$ 間の主効果が有意にみられた(条件間: $F(1, 52) = 10.87, p < .01$ ;  $\beta$ 間: $F(1, 52) = 47.46, p < .001$ )。

次に，言語的推定課題では，記述されたルールにおける，敵の数や逃げ道の数に関する記述の有無で参加者をコーディングした。そして，その度数について，記述の有無(2)×実験条件(4)のクロス集計表を作成した(表 1)。このクロス集計表について  $\chi^2$  乗検定を行った結果，敵についての記述の有無( $\chi^2(3) = 34.10, p < .01$ )，逃げ道の数についての記述の有無

( $\chi^2(3) = 26.26, p < .01$ )のどちらも有意差が見られた。残差分析の結果，敵の数についての記述は，敵条件と敵道条件で比率が高く，道条件で比率が低くなっていた。一方，逃げ道の数についての記述は，敵条件で低く，道条件で高かったが，敵道条件では比率の差は見られなかった。このことは，顕著性の高い情報も入力情報に使用している場合には，顕著性の低い情報も使用されにくくなることを示している。

これらの結果は，言語的推定課題において，顕著性の高い情報と低い情報を同時に利用する場合に，低い情報が行動ルールの推定に利用されにくくなることを示すものである。このことは，行動ルールの推定における言語隠蔽効果が見られる1つの証拠となるだろう。

また，この結果は，自律エージェントに対するユーザーの評価や理解度を調査する方法に対して示唆を与える。本研究の結果は，顕著性の低い情報が，高い情報とともに利用されている場合に，非言語的にはある程度利用しているにもかかわらず，言語的に表現させると無視されてしまうというものである。つまり，ユーザーの理解度を調査する場合に言語的な方法を利用してしまうと，ユーザーの理解度を過小評価してしまう可能性があるということである。自律エージェントに対する理解度を調査する方法を検討する際には，このような言語化の影響を加味する必要があるだろう。

表1 言語的推定課題における敵の数および逃げ道についての記述の有無についての度数と残差分析の結果( $p < .05$ の項目に差の方向を記述)

敵についての記述の有無		敵の数			
	敵条件	逃げ道条件	敵道条件	統制条件	
記述あり	27 ▲	10 ▼	24 ▲	23	
記述なし	0 ▼	17 ▲	2 ▼	8	

道についての記述の有無		道の数			
	敵条件	逃げ道条件	敵道条件	統制条件	
記述あり	1 ▼	19 ▲	13	14	
記述なし	26 ▲	8 ▼	13	14	

#### 4. 謝辞

本研究は，トヨタ自動車株式会社および科学研究費補助金 18H05320 の支援によって行われた。なお，本研究は著者ら独自の意見及び結論を反映したものである。

#### 参考文献

- [1]Rubin, N. (2001). Figure and ground in the brain. *Nature neuroscience*, 4(9), 857-858.
- [2]Wolff, P., & Song, G. (2003). Models of causation and the semantics of causal verbs. *Cognitive psychology*, 47(3), 276-332.
- [3]Kershaw, T. C., & Ohlsson, S. (2001). Training for insight: The case of the nine-dot problem. In *Proceedings of the Annual Meeting of the Cognitive Science Society* (Vol. 23, No. 23).
- [4]Schooler, J. W., Ohlsson, S., & Brooks, K. (1993). Thoughts beyond words: When language overshadows insight. *Journal of Experimental Psychology: General*, 122(2), 166–183.
- [5]Ball, L. J., Marsh, J. E., Litchfield, D., Cook, R. L., & Booth, N. (2015). When distraction helps: Evidence that concurrent articulation and irrelevant speech can facilitate insight problem solving. *Thinking & Reasoning*, 21(1), 76-96.
- [6] de Leeuw, J. R. (2015). jsPsych: A JavaScript library for creating behavioral experiments in a web browser. *Behavior Research Methods*, 47(1), 1-12. doi:10.3758/s13428-014-0458-y.