

## Premonitions of Insight Predict Impending Error

Janet Metcalfe (Indiana University)

*Journal of Experimental Psychology: Learning, Memory, and Cognition*, 1986, Vol. 12, No. 4, 623-634

- 本研究では、洞察問題やアナグラムにおける、正答・誤答に至るまでのダイナミックなメタ認知について調べる
  - 逐次的に暖かさ評定 (Warmth ratings) を測定して検討する
    - 「どのくらい解に近づいていると感じるか」
    - 鬼が隠した宝を子が探すゲームにおいて、子が宝に近づくと鬼は暖かく感じ、離れると冷たく感じる (Simon, Newell, & Shaw, 1979)
- 問題解決時の暖かさは、記憶再生時の TOT (Tip-of-the-Tongue) 現象に似ている (Brown & McNeill, 1966; Eysenck, 1979; Freedman & Landauer, 1966; Hart, 1967; Yarmey, 1973)
  - TOT や既知感 (Feeling of knowing) のようなメタ認知は、続く再生成績をうまく予測できる (Hart, 1965, 1967; Nelson, 1984; Nelson, Leonesio, Shimamura, Landwehr, & Narens, 1982)
- しかし、洞察問題の成績はメタ認知でうまく予測することができない (Metcalfe, 1986)
  - 解決前の暖かさ評定はほぼ一定で、解決した途端、急に跳ね上がる
- ここでは、暖かさ評定の正確さ (正誤との関係) について検討していく

### Experiment 1

#### Method

##### *Procedure.*

- 紀元前問題 (B.C. problem) が書かれたシートを配布
- 教示
  - 10 秒ごとに音になるので、音がなったら暖かさを評定
    - 0 (冷たい) ~ 10 (暖かい・解に至った) の 11 件法
  - 問題が解けたら、実験者に分かるようにその解を紙に記述
  - 制限時間は 5 分間だが、解を書き終えたらそこで終了

##### *Materials.*

A stranger approached a museum curator and offered him an ancient bronze coin. The coin had an

authentic appearance and was marked with the date 544 B.C. The curator had happily made acquisitions from suspicious sources before, but this time he promptly called the police and had the stranger arrested. Why?"

#### *Subjects.*

- 心理学入門の受講生 134 名
  - うち意味不明なことを書いた, または評定に欠損のある 36 名を分析から除外

#### **Results**

- 以後, 検定はすべて有意水準 5% で行う
- 解の内訳
  - 正しい解を記述した (正答) 43 名
  - 誤った解を記述した (誤答) 44 名
  - 解を記述しなかった (未答) 11 名

#### *Warmth as a function of correctness.*

- 正答者・誤答者のうち, 評定値 10 を出すまでに 3 区間以上評定を行った参加者 (正答者 19 名・誤答者 33 名) を以降の分析対象とする
- 解に至った区間の直前 3 区間の暖かさ評定値の平均値を算出した (Table 1 参照)
  - 正答  $\leq$  誤答 ( $F(1, 50) = 2.81, p = .09$ )
    - 直前 2 区間 (Second Last・Last) 正答  $<$  誤答 ( $F(1, 59) = 6.48$ )
    - 直前のみ (Last のみ) 正答  $<$  誤答 ( $F(1, 68) = 15.00$ )
  - 区間の主効果あり ( $F(2, 100) = 24.38$ )
    - 解に近づくにつれて暖かさが増す
  - 解の正誤と区間の交互作用なし

#### *Insight pattern analysis.*

- 暖かさ評定の推移のパターンを 2 つに分類
  - 最初の評定から最後の評定に至るまでの評定値の増加が…
    - 1 ポイント以下 「洞察型 (Insight pattern)」と定義<sup>1</sup>
    - 5 ポイント以上 「増加型 (Incremental pattern)」と定義<sup>2</sup>
- 正答・誤答それぞれにおける洞察型・増加型の割合を算出した (Table 2 参照)
  - 洞察型において, 正答 (.46)  $>$  誤答 (.14) ( $z = 2.78$ )

<sup>1</sup> 厳密には 0 ポイントとするべきだが, 多少ゆとりを持たせるため, 1 ポイントとした

<sup>2</sup> やや恣意的な定義だが, 最初の評定値が 1 程度なので, 十分起こり得る

- Table 1・2 では評定値の分布が分からないので、ヒストグラムを描いた (Figure 1 参照)
  - 誤答者は解 (実際には誤答) に近づくにつれて暖かさが増していく
    - 最頻値が徐々に右へ推移していく
  - 正答者にはそのような傾向が見られない
    - 最頻値がずっと 0 のまま

## Discussion

- 実験 1 の成果は主に 2 点
  1. 紀元前問題に正答した場合、暖かき評定は最低から最大に一気に跳躍する
    - ゆえに、紀元前問題は洞察問題に該当する
  2. 紀元前問題で、解が出る直前に暖かさを強く感じたら、誤答である可能性が高い

### *Insight.*

- 洞察問題を解くために必要な認知プロセスは、漸進的ではなく悉無律的であるだろう
- 逆に誤答者は、連続的近似方略 (Successive approximations strategy)・満足方略 (Satisficing strategy)・累積方略 (Accumulative strategy) を用いていたかもしれない
  - 満足方略 ある程度満足できる解を徐々に受け入れていく方略 (Simon, 1979)

### *Intelligence differences.*

- 洞察問題を正答するには、知能の高さや経験の豊富さが影響してくるだろうか？
  - 熟達者は正確な臨床評価を行うが、自信は他の人より低い (Oskamp, 1982) という先行研究と、実験 1 の結果が似ている

### *Demand for progress.*

- 誤答者には潜在的な時間のプレッシャーがかかっていたのかもしれない
  - 解を出すまでの区間数において、正答 (5.35) < 誤答 (8.08) ( $t(68) = 2.49$ )

### *Attention.*

- 正答者が暖かき評定を適当に行っていた可能性はないだろうか？
  - 評定値をひとつ前の値と同じにすることで、負荷の軽減を図ることができる (Kahneman, 1973; Shiffrn & Schneider, 1977)

## Experiment 2

- 実験 1 との相違は、1 参加者が行う問題を複数にすることで、参加者内比較を行う点

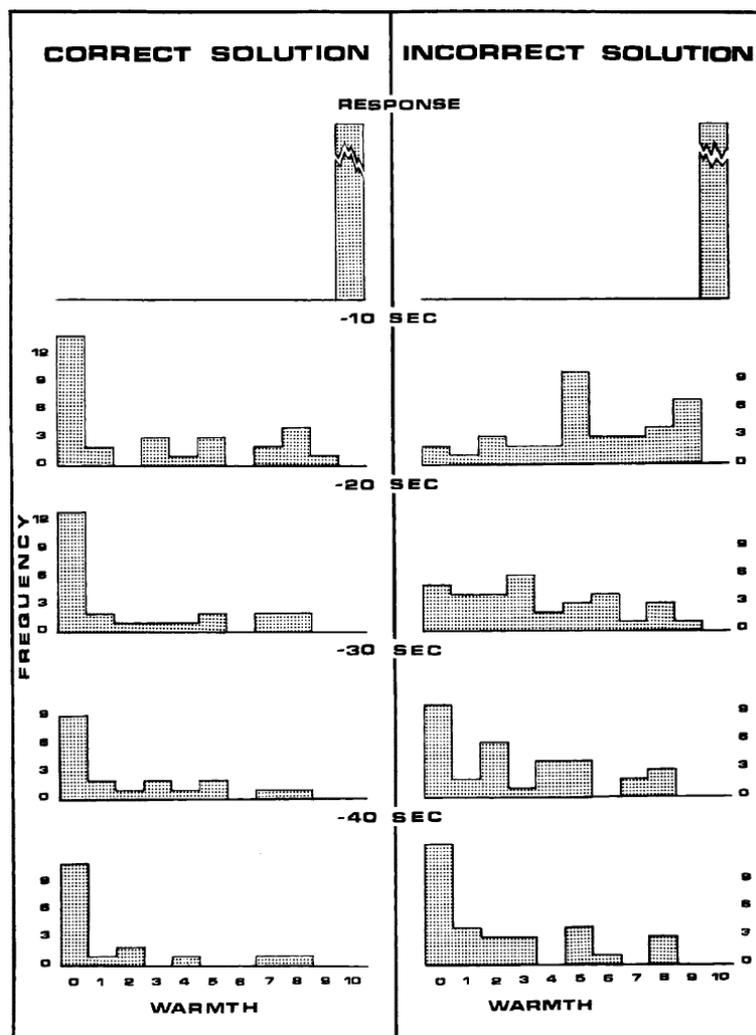


Figure 1. Distributions of warmth ratings for four intervals leading up to correct (left panel) and incorrect (right panel) responses to the B.C. problem, Experiment 1.

## Method

### Procedure.

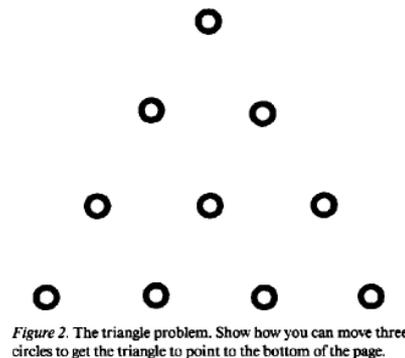
- 問題を 8 問提示
  - 3 問は即答できる易問, 5 問はすぐには解けない難問 (難問は以下詳細)
- 難問 5 問を難しいと思う順に並べる
- それぞれの難問に対して, 5 分間で解ける期待値を評定
  - 0 (確実に解けない) ~ 10 (確実に解ける)
- ランダムに提示された問題を解く
- 5 分間で解けなかった問題は再度シャッフルされ, 次の問題が提示される
- 15 秒ごとに 1 ~ 10 の 10 件法で暖かさを評定
- 暖かさ評定で 10 ポイントをつけたら, 解を紙に記述
- それを実験者が読み, 意味がよく分からなかったら再度明確に書くよう指示

*Subjects.*

- 心理学入門の受講生 44 名

*Materials.*

1. The horse trading problem: A man bought a horse for \$60 and sold it for \$70. Then he bought it back for \$80 and sold it for \$90. How much money did he make (or lose) in the horse trading business?
2. The chain problem: A woman has four pieces of chain. Each piece is made up of three links. She wants to join the pieces into a single closed ring of chain. To open a link costs 2 cents and to close a link costs 3 cents. She has only 15 cents. How does she do it?
3. The gardener's problem: A landscape gardener is given instructions to plant four special trees so that each one is exactly the same distance from each of the others. How would you arrange the trees?
4. The oil and vinegar problem: A small bowl of oil and a small bowl of vinegar are placed side by side. You take a spoonful of the oil and stir it casually into the vinegar. You then take a spoonful of this mixture and put it back into the bowl of oil. Which of the two bowls is more contaminated?
5. The postcard problem: Describe how to cut a hole big enough to put your head through in a (three by five inch) postcard.
6. The triangle problem (see Figure 2): The triangle points to the top of the page. Show how you can move three circles to get the triangle to point to the bottom of the page.



**Results**

- 難問の解の内訳 正答.28, 誤答.49, 未答.23
- 分析対象
  - 3 区間以上評定を行い正答した問題が 1 問以上ある, かつ 3 区間以上評定を行い誤答した問題が 1 問以上ある参加者 29 名

*Warmth as a function of correctness.*

- 解に至った区間の直前 3 区間の暖かさ評定値の平均値を算出した (Table 1 参照)
  - 正答 (3.95) < 誤答 (4.61) ( $F(1, 28) = 4.80$ )
  - 区間の主効果あり ( $F(2, 56) = 18.37$ )
  - 解の正誤と区間の交互作用なし

*Insight pattern analysis.*

- 正答・誤答それぞれにおける洞察型・増加型の割合を算出した (Table 2 参照)
  - 洞察型において, 正答 (.66)  $\div$  誤答 (.56) ( $z=1.2$ )
    - 期待値評定を行い, 実験 1 より暖かさ評定の初期値が高くなったため?

*Time.*

- 解に到達するまでの平均区間数
  - 正答 (9.33) > 誤答 (6.56) ( $t(28) = 2.40$ )
    - 実験 1 とは逆の傾向

*Expectancies.*

- 期待値評定の中央値で期待値高群と低群に分ける
  - 解の直前 3 区間における暖かさ評定値において,  
期待値高群 (5.23) > 低群 (3.22) ( $F(1, 27) = 6.84$ )
  - 正答割合において, 期待値高群 (.30)  $\div$  低群 (.33) ( $t(27) = .61, p = .55$ )

*Discussion*

- 実験 1 と同様, 直前においては, 誤答のほうが正答よりも強く暖かさを感じていた
  - これは解に至るまでに要した時間が長く, プレッシャーがあったためではない
    - むしろ正答のほうが長かった
  - また, 知能や経験などの個人差によるものではない
    - 参加者内比較を行っても同様の結果であったため
- 正答と誤答で暖かさが異なるのは, 洞察問題がミスリードさせる性質を持つため
  - 正答よりミスリードされた誤答に対してより暖かさを感じるためだろう
- では, 意図的なミスリードのないアナグラムのような問題ならどうだろうか?

### Experiment 3

*Method*

*Materials.*

- アナグラム 20 問 (Sherman, personal communication, 1985)

ssoia phmny aeuvl pmuoi ttnua rdcei oapnr tlcee reckl elcsa piaot ulipp nitga ocbna ucrco dtuai mnaai oocnl aebrl augdr
--

### *Procedure.*

- アナグラムを 1 ページに 1 問ずつ提示
  - 解く際にメモは使用できない
  - 制限時間は 1 問につき 5 分（解を記述したらその時点で次ページへ）
- 10 秒ごとに 1 ～ 10 の 10 件法で暖かさを評定
  - 次のページに移ったら、解き始める前に暖かさ評定 1 をつける
- 解を思い付いたら紙に記述して、暖かさを評定<sup>3</sup>

### *Subjects.*

- 心理学入門の受講生 24 名（うち分析対象 16 名）

### *Results*

- 解の内訳 正答 74%，誤答 7%，未答 19%
- Figure 4
  - 正答は 40 秒前からずっと最頻値が 1 だが、最後だけ急激に暖かさが増す
  - 誤答も最初は正答とよく似ているが、解に至った時点では大きく異なる
  - 解の時点における暖かさ評定において、正答 (9.78) > 誤答 (7.86) ( $t(16) = 3.92$ )
- Table 1
  - 正答 ≍ 誤答
  - 区間の主効果あり ( $F(2, 30) = 5.84$ )
  - 解の正誤と区間の交互作用なし
- Table 2
  - 洞察型において、正答 (.79) ≧ 誤答 (.66) ( $z = 1.63$ )

### *Discussion*

- 実験 3 の暖かさ評定は誤答に対する不安さを反映していたのだろう
  - 実験 1・2 とは異なり、必ずしも評定値 10 をつける必要がなかった
- 誤答ながらも高く評定した参加者は満足方略を使用していたかもしれない
  - 必ず評定値 10 をつけるようにすれば、さらに誤答時の暖かさが増すのでは？

## **Experiment 4**

<sup>3</sup> 「解が分かったら暖かさ評定は必ず 10 にしてください」と伝え忘れた（手続き上のミス）

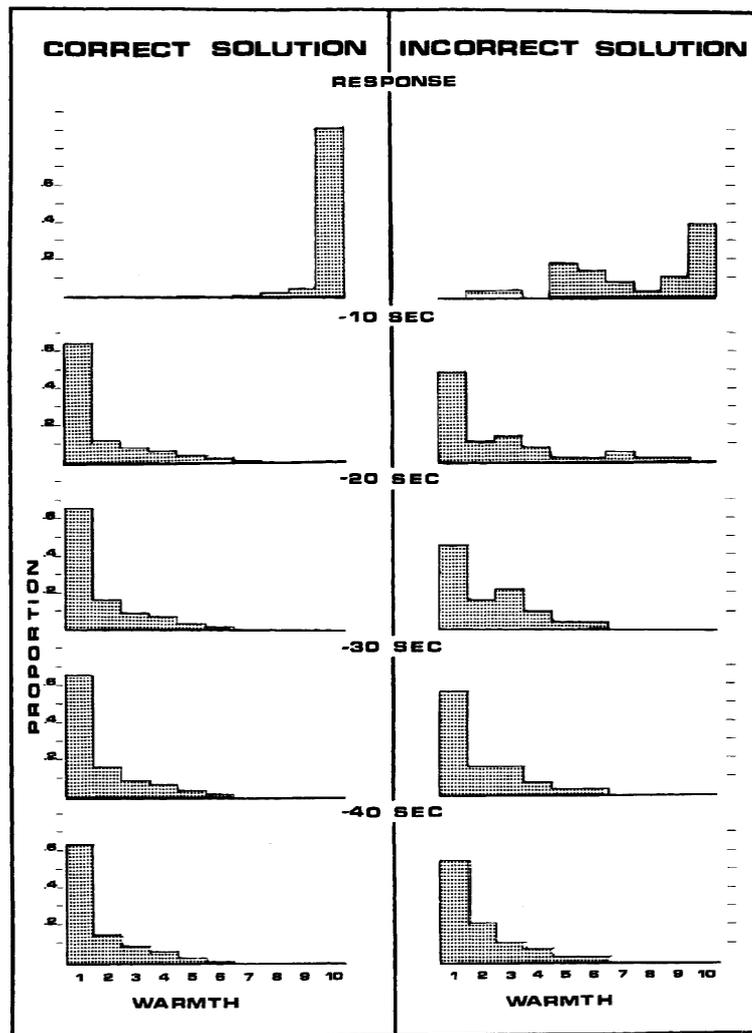


Figure 4. Distributions of warmth ratings for four intervals leading up to correct (left panel) and incorrect (right panel) responses to anagrams, Experiment 3.

#### Method

- 以下の点を除き，実験 3 と同じ
  - 暖かき評定を 0 ～ 10 の 11 件法に
  - 暖かき評定に 10 をつけてから解を記述するよう教示
- 心理学専攻の 2 年生，または張り紙で募集した参加者計 41 名（うち分析対象 23 名）

#### Results and Discussion

- Table 1
  - 正答 (1.66) ≤ 誤答 (2.11) ( $F(1, 22) = 4.0, p = .055$ )
    - 自信が十分ついてから解を記述するよう求めると，誤答の評定値が高くなる
  - 区間の主効果あり ( $F(2, 44) = 13.82$ )
  - 解の正誤と区間の交互作用なし

- Table 2
  - 洞察型において、正答 (.76)  $\approx$  誤答 (.76) ( $z = 1.63$ )

### Experiment 5

- 実験 3・4 から、誤答を導いたときには、満足方略や自信を納得させてしまう方略を使用していることが多々あるようだ
  - これらの方略は推測を許すことでさらに使用されるのでは？

### Method

- 以下の点を除き、実験 4 と同じ
  - 「解には普段見慣れない単語もあるので、どんどん推測してください」と強調
- 受講生と張り紙で募集した参加者 24 名（うち分析対象 20 名）

### Results

- Table 1
  - 正答 (2.33) < 誤答 (3.03) ( $F(1, 19) = 6.85$ )
  - 区間の主効果あり ( $F(2, 38) = 18.19$ )
  - 解の正誤と区間の交互作用なし
- Table 2
  - 洞察型において、正答 (.78) > 誤答 (.44) ( $z = 5.93$ )
    - 推測を促されると、正答時に洞察型が現れやすくなる

### Comparison of Experiment 4 and 5

- 実験 4・5 の違いは、推測を許したか（推測許可：実験 5）否か（統制：実験 4）
  - 洞察型の出現割合
    - 正答 統制  $\approx$  推測許可
    - 誤答 統制 (.76) > 推測許可 (.44) ( $z = 3.88$ )
  - 解の割合
    - 正答 統制 (.64) > 推測許可 (.55)
    - 誤答 統制 (.09) < 推測許可 (.27)
    - 未答 統制 (.27) > 推測許可 (.18)
  - 暖かさ評定値
    - 統制 (1.89) < 推測許可 (2.68) ( $F(1, 41) = 2.22$ )
- 推測を許可されると、満足方略を使用しやすくなる

**Additional Analyses**

- 暖かさ評定における正答と未答の比較 (Table 3 参照)

**Table 3**  
*Temporal Comparison of Warmth Ratings for Correctly Versus Unsolved Problems or Anagrams in Experiments 2, 3, 4, and 5*

Type of problem	Quarters			
	1	2	3	4
<b>Experiment 2 (problems)</b>				
Correct	3.69	3.86	3.96	4.53
No solution	3.04	3.19	3.05	2.84
<b>Experiment 3 (anagrams)</b>				
Correct	1.65	1.87	2.01	2.13
No solution	1.70	1.91	2.10	2.21
<b>Experiment 4 (anagrams, no guessing)</b>				
Correct	1.43	1.70	1.68	1.61
No solution	1.36	1.47	1.51	1.43
<b>Experiment 5 (anagrams, guessing)</b>				
Correct	1.54	1.64	1.87	2.45
No solution	1.26	1.50	1.49	1.52

- 部分的に、正答 > 未答
  - 正答時の評定は床に張りついているわけではない (適当にやっているのではない)

**Conclusion**

- 問題解決時にどのようなプロセスをたどるかによって、解けるか否かが変わってくる

	正答	誤答
洞察型	.76	.52
増加型	.12	.25

- 洞察は飛躍的であるにもかかわらず、直前3区間で多少暖かさが増すのはなぜか？
  - 大半の参加者が最小値から評定を始めるので、減ることがあまりないため
  - すでに洞察が起こっているが、それを明文化できない状態が暖かく感じる？
- 誤答を出す直前のほうが、正答を出す直前よりも暖かく感じていた
  - これは、誤答を出すまでの時間のほうが長いからではない (区間数に差なし)
  - また、個人差によるものでもない (参加者内比較でも有意)

- 加えて、洞察問題の特徴で説明できるものでもない（アナグラムでも起こる）
- さらに、正答時には評定を怠っていたわけでもない（未答のほうがもっと低い）
  
- 誤答時には、エレガントではない解であっても自身を納得させてしまう満足方略が用いられている
  - この方略は推測を許されると一層使用されやすくなる
- ただ、満足方略は現実場面であれば有効な方略であろう
  
- 本研究の結果から、洞察の兆候を調べれば、誤答をある程度予測できると言えるかもしれない

**Table 1**  
*Warmth Ratings on the Last Three Intervals Before Solution for Correct and Incorrect Responses in Experiments 1, 2, 3, 4, and 5*

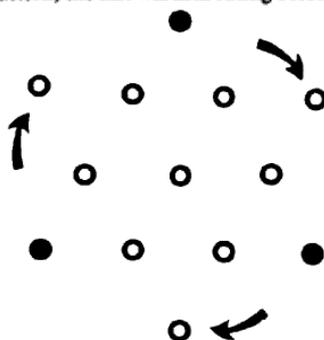
Response	N	Interval rating			Solution
		Third last	Second last	Last	
Experiment 1 (problem)					
Correct	19	2.05	2.42	3.47	10
Incorrect	33	2.92	3.57	5.25	10
Experiment 2 (problems)					
Correct	29	3.52	3.73	4.60	10
Incorrect	29	4.16	4.64	5.02	10
Experiment 3 (anagrams—10 not required)					
Correct	16	2.06	2.19	2.62	9.78
Incorrect	16	2.10	2.16	2.53	7.86
Experiment 4 (anagrams—no guessing)					
Correct	23	1.29	1.50	2.20	10
Incorrect	23	1.78	1.91	2.63	10
Experiment 5 (anagrams—guessing)					
Correct	20	1.92	2.25	2.81	10
Incorrect	20	2.61	2.67	3.83	10

*Note.* In Experiment 1, subjects are divided into two groups on the basis of their performance on the one problem; in the other experiments, subjects' correct and incorrect responses on all problems are considered.

**Table 2**  
*Proportion of Problems or Anagrams Showing an Insight or an Incremental Pattern of Warmth Ratings*

Condition	Insight pattern		Incremental pattern		N	
	Correct	Incorrect	Correct	Incorrect	Correct	Incorrect
Experiment 1	.46	.14	.21	.62	24	37
Experiment 2	.66	.56	.11	.08	56	80
Experiment 3	.79	.66	.06	.06	227	32
Experiment 4	.76	.76	.15	.22	385	55
Experiment 5	.78	.44	.14	.32	185	108
Overall	.76	.52	.12	.25	877	312

*Note.* For the insight pattern, the last warmth rating before the rating with which the response was given was no more than 1 point greater than the first rating given in the protocol. For the incremental pattern, the last warmth rating before the rating with which the response was given was at least 5 points greater than the first rating given.



*Figure 3.* The triangle problem solution. The three points of the triangle are rotated around the central rosette.