

Effects of Hypothesis Generation on Hypothesis Testing in Rule-Discovery Tasks

Dennis. J. Adsit and Manuel London

The Journal of General Psychology, 1997, 124(1), 19-34

- はじめに
 - ▶ 認知的仮説検証(cognitive hypothesis testing)は不良定義問題の不確実性を減らす過程である(Abelson & Levi, 1985)
 - ▶ 問題状況における対象を支配するルールの発見も含む

- Hypothesis-Testing Research
 - ▶ 問題解決者は現在の仮説を確認したがる(Bruner, Goodnow, & Austin, 1956)
 - ▶ 最終的な仮説の証明は不可能なので、反証を探す必要がある(Popper, 1959)
 - ▶ 2-4-6 課題(Wason, 1960)による検討
 - ・ 2-4-6 の例から初めて、仮説を調整して、正しいルールを見つける
 - ・ 探すべきルールは単純な「小さい順の 3 つの数字」だが、典型的にたてられる最初の仮説は「連続する偶数」や「2 ずつ増える数」である
 - ・ アイデアを確認しようとするために問題解決エラーを起こす(Tversky & Kahneman, 1974)
 - ▶ いくつかの初期の例と異なる結果もある
 - ・ 最初の 3 つの数字を変えると正しく解ける(Wetherick, 1962)
 - ・ 反証仮説を示すとパフォーマンスが向上(Gorman & Gorman, 1984)
 - テストされた仮説の情報を、ほかの仮説を考えたり、考えなかったりしながら対立仮説を消すのに利用している(Tukey, 1986)
 - ▶ ほかのルール発見課題
 - ・ 課題複雑性が高いと新しい仮説を生み出すことを困難にさせる (Mynatt et al., 1978)
 - ・ 仮説を証明するというゴールが対立仮説の検討を妨げる(Dunber, 1993)
 - ▶ 確証バイアスではなく、判断上有用な情報を評価できない(Fischhoff & Beyth-Marom, 1983)
 - ▶ 確証的な仮説検証が現実的な条件ではよい方法の場合もある(Klayman & Ha, 1987)
 - ・ 最初の仮説とターゲット仮説の 3 つの関係性が重要
 - 反証方略が有効
 - embedded:最初の仮説の全要素はターゲット仮説の要素

- 確証方略が有効

overlapping: ターゲット仮説と共通な要素とそうでない要素がある

surrounding: ターゲット仮説の全要素が最初の仮説の要素のサブセット

- ▶ 仮説検証は動的なプロセスなので課題全体における仮説検証の必要がある (Hogarth, 1981; Lopes, 1986)

● Hypothesis-Generation Research

- ▶ 仮説生成の影響を分離させた研究はあまりない
- ▶ 仮説生成の不足が仮説検証の成功を制限するのかを調べる必要がある
 - ・ 仮説生成とは与えられた情報から可能な、対立仮説を作り出すこと (Fisher, Gettys, Manning, Mehle, & Baca, 1983)
- ▶ 質と量において“乏しい”仮説のセットを作り、その完全性と妥当性を過大評価してしまう (Fisher et al., 1983; Gettys, Fisher, & Mehle, 1978; Getty, Manning, Melhe, & Fisher, 1980; Mehle, Gettys, Manning, Baca, & Fisher, 1981)
- ▶ 仮説検証の前に仮説を生成させる
 - ・ 実験のデザイン前に仮説を考える時間をとることにより課題を正確で早く解決できる (Klahr & Dunbar, 1988)
 - ・ 確証バイアスは仮説生成の機能 (function) である (Farris & Revlin, 1989)
 - 仮説の選択に反事実推測方略 (counterfactual reasoning strategy) を使い、各仮説と対立仮説を比較する
- ▶ 実際には成功を高く予測できる強い推論を行っている
 - ・ 仮説の評価のみを含む課題ではあらかじめ指定した課題のセットから一つ一つばらばらに評価する
- ▶ 自分で仮説を作ると正しさに自信がなく、仮説の正確さに敏感 (Koehler, 1994)
 - ・ 仮説生成は仮説の評価だけよりも対立仮説のテストを行いやすい

● Current Research

- ▶ 4条件デザイン
 - ・ baseline: 課題を与えられて正しいルールを見つける
 - ・ participant-generated: 検証を行う前に仮説のリストを作る
 - ・ yoked: 別の被験者の作った仮説のリストを仮説検証の前に与える
 - ・ experimenter-supplied: ターゲットルールを含む仮説のリストを与えられる
正しい解決がリストの中にあることを告げられる
- ▶ participant-generated より
 - 生成する仮説の質や量と仮説検証のパフォーマンスには相関がある
- ▶ participant-generated と yoked の比較より

- 自分で仮説を生成すると追加的な仮説や対立仮説が考えにくい
- ▶ 確証バイアスにより仮説検証がうまくいかなかったと結論付けるのは難しい
 - ・ 被験者は単なる仮説検証以上のこと（仮説生成）を行っている
 - ・ 正解、不正解のレベル以上の微妙な指標を見る必要がある
- ▶ experimenter-supplied より
 - ・ 仮説検証を仮説生成から独立して調べることができる
 - 反証テストを行い正しいルールを見つけることができる
- ▶ 正しいルールの発見の予測
baseline < participant-generated < yoked < experimenter-supplied
- ▶ 追加的な 2 つの指標
 - ・ 仮説検証に費やした時間
 - 正しいルールが発見できる順に仮説検証が短いだろう
 - ・ 先の仮説テストで棄却された仮説を保持、再テストした程度
 - テストの前に仮説生成を行うと仮説の保持が減る傾向がある(Klahr & Dunbar, 1988)
 - リストを持つ 3 条件と持たない条件で仮説の保持時間が異なるだろう
 - ・ 発見過程で行った検証
 - 反証と対立テストの割合はルールの発見と同じ順になるだろう
- ▶ 課題の性質も影響するため、3 つの異なる課題とそれぞれに 3 つの異なるターゲットルールを用いる

● Method

□ participants

- ▶ 108 名の学生：男性 61 名 女性 47 名
- ▶ 18～46 歳 平均 20.6 歳

□ design

- ▶ 反復測定 Greco-Roman Latin square デザイン
 - ・ 3 つの課題 (A,B,C) とルール (1,2,3) のシーケンスを設定する
 - a) A1-C3-B2 b) B3-A2-C1 c) C2-B1-A3
 - ・ 同じ仮説条件で 3 つの課題と 3 つのルールをすべて行う
 - ・ a)、b)、c)を 1 つの反復測定とみなす（実際には 3 人の被験者が行う）

□ procedure

1. task and testing procedure

- a) 2-4-6 課題(Wason, 1960)

- ・ 実験者の考えるルールを当てる
- ・ 最初は「2-4-6」からはじめる
- b) 都市課題(Klayman & Ha, 1989 に類似)
 - ・ 3つの都市のセットを描写するルールを見つける
 - ・ 世界地図にカラークリップで3つの都市を示す
 - ・ 最初は南アメリカの「Brasilia, Asuncion, Cordoba」からはじめる
- c) 広告課題(Bruner et al., 1956 に類似)
 - ・ 広報ディレクターになったつもりで新商品の広告をデザインする
 - ・ 広告はメディア、ポジション、伝播スタイル、エモーショナルアピールからなる
 - ・ 実験者の考える成功広告のキーを考え、各カテゴリーの要素を決める
 - ・ ひとつの成功広告例が提示される
- ▶ 全課題レコードシートに以下の3つを記入後、正しいルールに一致しているかどうかを示される
 - ① 実験者のルール（広告要素のキー）を予測
 - ② テストしたい3つの数字（都市、広告）を提示
 - ③ テストしたい理由
- ▶ 最後に最終的な結論を記録し、レコードシートを回収、次の課題に移る
- ▶ 18 試行まで、または 30 分まで行う
- ▶ 最終推論を示せばやめられる（フィードバック無し）

2. rule

- ▶ 最初の仮説とターゲットルールの関係が **embedded, overlapping, surrounding** になるようにルールを設定
 - a) 2-4-6 課題の場合
 - ・ 典型的な最初の仮説「偶数」「2 ずつ増える数」
 - **embedded**: 順番に増える数字
 - **overlapping**: 一桁の数字
 - **surrounding**: 連続的に増える偶数
 - b) 都市課題
 - ・ 典型的な最初の仮説「南アメリカの都市」
 - **embedded**: 北、中央、南アメリカ大陸の都市
 - **overlapping**: 南半球の都市
 - **surrounding**: 赤道より下の南アメリカの都市
 - c) 広告課題
 - ・ 典型的な最初の仮説はないためターゲットセットのサイズでルールを設

定

- 大:プリントメディアを使った広告
- 中:humor/fun を使った広告
- 小:humor を使った一般誌広告 (成功広告を新しく 3 種類しか作れない)

3. Hypothesis-generation conditions

- ▶ **baseline:** 特別な手続きはない
- ▶ **participant-generated:** 例と一致するできるだけ多くのルールを考え、5 分間でシートに記入させる
- ▶ **yoked: participant-generated** 条件の学生が考えたルールのリストを与えられ、最初の試行前に 5 分間見せる
※ルールは別の参加者が作ったものであること、正しいかどうか、その参加者が解決できたかはわからないと告げる
- ▶ **experimenter-supplied:**正しいルールを含む 9 つのルールを与えられる
※1 つは正しいルールであることを告げる
※ターゲットルールにかかわらず同じ 9 つのルールを与えられる
※6 つはターゲットセットを広く覆うルール、最初の例にはすべて当てはまる

4. Dependent measures

- ▶ 正しいルールを発見できたか
- ▶ 解決までの時間、試行数
- ▶ レコードシート (反証、確証、対立仮説の区別)
- ▶ 棄却された仮説の保持の程度

□ analyses

- ▶ 3 タイプの試行は割合に換算
- ▶ **whole-plot** と **split-plot** の組み合わせ - 36 の 3×3 の square を用いる
 - ・ **whole-plot**…各条件間の比較
 - ・ **split-plot**…36 の square それぞれの比較
 - 課題の性質と条件間に交互作用があるか
 - ・ 課題、ルール、シークエンスの主効果と条件との交互作用も調べる

● Result

- ▶ 4 条件の推論の正しさに有意差有り ($F(3, 324) = 9.09, p < .001$)
計画比較(planned comparison): experimenter-supplied > 残りの 3 条件

$(t(320) > 4.55, p < .001)$

⇒仮説の生成や仮説を受け取ることはパフォーマンスに影響しない

- ▶ participant-generated における分析
 - ・ 平均仮説生成数 8.3 (SD=3.0)
 - ・ 課題間の平均仮説生成数に有意差なし (課題ごとに可能な仮説の数は異なる)
 - ・ 5分以内のターゲット仮説の生成は 20/81 パフォーマンス
 - ・ 仮説生成数と正しいパフォーマンスの間の相関は .03
 - ⇒多くの仮説を生成しても課題が正しく解けるとは限らない
 - ・ 仮説検証の前にターゲット仮説を生成したかと正しい解決の間には関係有り
 - $(\chi^2 = 4.02, p < .05, \phi = .25)$
- ▶ 仮説の保持は全条件においてめったに起こらない
- ▶ 仮説の生成やリストを与えられることは仮説検証のパフォーマンスに促進効果を持つ
 - ・ 仮説検証にかかった時間
 - baseline > participant-generated, experimenter-supplied, yoked
 - (all $t_s \geq 3.13, p < .005$)
 - ※仮説生成の5分を加えると baseline, participant-generated, yoked の有意差はなくなる
- ▶ 反証試行の割合…主効果有り ($F(3, 24) = 6.46, p < .001$)
 - experimenter-supplied > 残りの3条件 ($t(320) > 5.75, p < .01$)
- ▶ 対立仮説の割合…主効果無し ($F(3, 24) = 1.10, p > .05$)
- ▶ 反復の主効果無し ($F(8, 24) < 1.0$)
- ▶ 仮説生成条件と課題、ルール、シーケンスの正答率における有意差なし
 - $(F(6, 192) < 1.52, p > .05)$
 - ⇒変数の効果は課題とルールの間で一般化される
- ▶ 正答割合における課題間の主効果有り ($F(2, 192) = 10.76, p < .001$)
 - Tukey の HSD : 2-4-6 課題 < 広告課題 ($p < .05$)
 - ⇒広告課題が一番簡単
- ▶ ルールの主効果有り ($F(2, 192) = 5.23, p < .01$)
 - Tukey の HSD : ルール 1,2 > ルール 3 ($p < .05$)
 - ⇒最も制限のあるルールが解決されにくい
- ▶ 順序の主効果 ($F(2, 192) = 3.53, p < .05$)
 - Tukey の HSD : 有意差無し ($p < .05$)

● Discussion

- ▶ 仮説検証の前に仮説のリストを生成することはルール発見に影響しない

- ▶ 仮説のソースはルール発見パフォーマンスに影響しない
- ▶ **experimenter-supplied** が最も反証的で正答率が高かった
 - ・ 正しいルールが残るまでリストからルールを消去していった
 - ⇒反証が増えた（必ずしも反証を使う必要はない）
 - ・ **participant-generated** や **yoked** はリストを見ないことも珍しくなかった
- ▶ 多くの仮説を生成してもルール発見の可能性は上がらない
- ▶ ターゲット仮説を生成することは正しい解決と正の関係がある
 - ・ 被験者の仮説生成の乏しい能力が仮説検証パフォーマンスの障害となる
- ▶ 最も制限されたルールがもっとも解決されにくい
- ▶ 仮説保持に条件間の差はない
- ▶ 一般化可能な結果に興味があるなら仮説検証課題では仮説検証を調べるべき
 - ・ 特に最初の仮説とターゲットルールとの関係を調べるべき
- ▶ 仮説生成と仮説検証の関係をさらに検討すべきである
- ▶ 反証は仮説検証の効果に重要な役割を果たす

TABLE 1
Means and Standard Deviations for Percentage Correct, Confirmations, Disconfirmations, and Alternative Tests

Variable	Percentage correct		Percentage confirmations		Percentage disconfirmations		Percentage alternatives tests	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
Hypothesis-generation conditions								
Baseline	.42	.50	.64	.33	.18	.21	.18	.25
Subgenerated	.53	.50	.63	.34	.18	.22	.19	.25
Experimenter-supplied	.86	.35	.38	.32	.40	.31	.22	.26
Yoked	.53	.50	.56	.36	.19	.20	.25	.26
Task								
2-4-6	.46	.50	.60	.35	.22	.25	.18	.24
Cities	.57	.50	.55	.33	.27	.25	.17	.20
Advertising	.72	.45	.50	.37	.22	.26	.28	.30
Rule								
Embedded	.64	.48	.54	.34	.26	.25	.20	.24
Overlaps	.64	.48	.53	.37	.22	.25	.24	.28
Surrounds	.48	.50	.58	.35	.22	.26	.19	.24
Sequence								
1	.66	.47	.46	.35	.28	.26	.26	.26
2	.51	.50	.60	.34	.20	.24	.19	.23
3	.59	.49	.60	.36	.22	.26	.18	.27

TABLE 2
Analyses of Variance

Source	<i>df</i>	Percentage correct		Percentage disconfirmations		Percentage alternatives tested	
		<i>SS</i>	<i>F</i>	<i>SS</i>	<i>F</i>	<i>SS</i>	<i>F</i>
Conditions	3	9.00	9.09***	.91	6.46***	.32	1.10
Replications	8	1.25	.48	.54	.47	.75	.90
Conditions × Replications	24	7.89		3.66		2.36	
Tasks	2	3.65	10.76***	.20	3.16*	.82	10.25***
Rules	2	1.78	5.23**	.12	2.00	.15	2.00
Sequences	2	1.19	3.53*	.36	6.00**	.42	5.25**
Conditions × Tasks	6	1.18	1.17	.16	1.00	.36	1.50
Conditions × Rules	6	.76	.76	.21	1.33	.60	2.50*
Conditions × Sequences	6	1.57	1.52	.43	2.33*	.67	2.75**
Participants (within)	72	17.78	1.47	6.30	3.00***	7.14	2.50*
Error	192	32.53		6.04		7.52	
Total	323	78.58		20.93		21.11	

* $p < .05$. ** $p < .01$. *** $p < .005$.