

The Cognitive Reflection Test as a predictor of performance on heuristics-and-biases tasks.

Maggie E. Toplak & Richard F. West & Keith E. Stanovich, 2011, Mem Cogn, 39, 1275-1289.

1. Introduction

- 認知内省テスト(The Cognitive Reflection Test ; CRT)は 3 項目の尺度(Frederick, 2005)
 - (a) バットとボールは合計で 1.10 ドルします。バットはボールより 1 ドル高いです。ボールは何セントですか？(A.5 セント)
 - (b) 5 つの機械ならば 5 分で 5 つの製品を作れます。100 の機械で 100 の製品を作るには、何分かかりますか？(A.5 分)
 - (c) 湖にスイレンの土地があります。毎日土地は 2 倍になります。もし湖全体を覆うのに 48 日かかるならば、湖の半分を覆うのは何日ですか？(A.47 日)
- 有力な誤った回答の選択肢を覆す傾向を測り、正しい反応を導くようにさらに内省する
 - 特徴：早く、直観的に答えが頭に浮かぶが、早い答えは誤っている
 - 正しい結果を導き出す鍵は頭に浮かんだ最初の結果を抑えるもしくは評価すること
- CRT はヒューリスティックバイアス課題の予測因子として理想的な構成のように見えるが、先行研究は一貫していない
- 本論文では、より広範囲のヒューリスティックバイアス課題における CRT の予測特性を調査
- CRT のいくつかの潜在する精神的な構造を明らかにすることを試みた
- 知能テストや認知態度テストは、思考傾向の尺度が典型的なパフォーマンス状態で評価されているかどうかについて、最も有効なパフォーマンス評価
 - CRT は認知能力と思考傾向を利用する事実から、その予測因子として可能性がある
- 本研究では、広範囲のヒューリスティックバイアス課題と意思決定課題によって、CRT の予測分散を分離することを試みた
 - さらに三段論法の信念バイアスの程度を予測する能力を実験した
- CRT がそれらの合理的思考尺度と共有する変数が、さらに認知能力と思考傾向によって共有されるかどうかを調査した
- CRT がヒューリスティックバイアス課題を予測する原因を調査した研究で用いられた認知能力尺度と思考傾向を補足するために、実行機能課題を加えた

2. Method

2.1. Participants and procedure

- 参加者 346 名(男性 95 名、女性 251 名；平均年齢 = 20.1, $SD = 3.9$)
 - ほとんどが大学生
- 所要時間：2 時間

2.2. Tasks and variables

2.2.1. Cognitive reflection test(Frederick, 2005)

- 認知内省テスト 3 項目
- 分析には 3 項目のパフォーマンスの複合測定量を使用
- 正しい結果の平均パフォーマンスは 0.7 項目 ($SD = 0.93$)
 - 間違えた参加者：55.8%(193 名)
 - 3 項目とも正しい参加者：6.6%(23 名)

2.2.2. Cognitive ability(WASI; Wechsler, 1999)

- 語彙テスト、行列推論テスト；言語非言語能力の指標として使用
 - 語彙テストの平均素点：52.6($SD = 7.4$)
 - 行列推論テストの平均素点：27.3($SD = 3.7$)
- 語彙テスト、行列推論テストの素点は z スコアに変換され、認知能力の複合測定量を作成するために合計された

2.2.3. Heuristics-and-biases tasks

- 合理的思考の重要な特徴を反映した 15 個のヒューリスティックバイアス課題
 - ベースレート問題／サンプルサイズ問題(2 つ)／感受性を評価する問題／ギャンブラーの誤り問題(2 つ)／結合問題／共変動発見問題／方法論的推論問題／ベイジアン推論問題／フレーム問題／基準無視を評価する問題／判断一致可能性／サンク・コスト問題／結果バイアス問題
- 0or1 得点で複合得点を算出($M = 6.88$, $SD = 2.32$)

2.2.4. Syllogistic reasoning problems with belief bias

- 結論の信頼性と論理の妥当性が対立する演算的推論項目 3 つ
 - 全ての生物は水が必要だ。バラには水が必要だ。バラは生物だ
 - 結果は前提に従っているか否かを回答
- 参加者が肯定式の演算的确实性を認めるか否かを評価する演算的推論課題 2 つ
 - もし車がホンダなら、それは高価だ。ジョンの車はホンダだ。ジョンの車は高価だ。
 - 真実か否かを true ~ false の 7 段階で評定(真実を 1、その他を 0 として分析)
- 5 つの推論問題全体で正解の平均値：2.72($SD = 1.21$)

2.2.5. Executive-functioning measures

2.2.5.1. Set shifting

- The Trail making Test (Reitan, 1955, 1958)
 - 1→A→2→B→3→C→…→L→13 と繋げていく
- 平均時間：59.6s($SD = 24.6s$)
- 平方根変換した後、 z 得点に変換された
 - より高い得点がより高い set-shifting 能力(強化学習の状況の変化に直面した際の柔軟さ)を示すように z 得点は反映された

2.2.5.2. Inhibition

- ストループ課題(The Stroop task)
- 3 条件それぞれで 24 項目(4×6)

- word-reading 条件、color-naming 条件、衝突条件
- 独立変数：衝突条件の総回答時間から色条件の総回答時間を引いたもの
- 平均時間：10.2s($SD=5.0$, 範囲：0.1~27.1s)
- より高い得点がより抑制能力があることを示すように z 得点は反映された

2.2.5.3. Working memory

- 定速聴覚連続付加検査(The Paced Auditory Serial Addition Test(PASAT; Gronwall, 1977))
 - ワーキングメモリ、分割的注意、情報処理速度を評価する課題
- 連続で提示される一桁の数字を即座に足し算する
 - Trial1 : 3s、Trial2 : 2s
- 分析には、それぞれの Trial で正しく足し算された合計値を用いた
- Trial1 と Trial2 の平均得点：38.3($SD=9.3$)
- ※ ワーキングメモリは実行機能尺度よりも認知能力尺度と高い相関があるので、ワーキングメモリを基にもう一つの認知能力指標(CA2)を作成
 - WASI の合成得点とワーキングメモリ課題を合計

2.2.6. Thinking dispositions

- 主観報告アンケート
- それぞれの項目に同意するか、6段階評定
- 1つめ：Actively Openminded Thinking scale ; 41項目
 - より高い得点はよりオープンマインド思考へ向かう傾向を示す
ex. 「人々はいつも彼らの信念に反する証拠を考慮に入れるべきである」
 - 41項目の合計得点の平均値：161.1($SD=19.6$)
- 2つめ：迷信を信じる思考の尺度 ; 13項目
 - 超常現象尺度(2項目)、幸運尺度(4項目)、ESP 尺度(4項目)、迷信思考尺度(3項目)を合成
ex. 「13 は不幸の数字だ」
 - 13項目の合計得点の平均値：33.5($SD=10.4$)
- 3つめ：Consideration of Future Consequences (CFC) ; 12項目
 - 行動を選ぶ際に個人が遠い結果を考慮する範囲を計測
ex. 「私は時間が解決するだろうと考え、目の前の問題を解決するためだけに行動する」(逆転項目)
 - 12項目の合計得点の平均値：48.2($SD=7.4$)

3. Results

3.1. Table1 : ヒューリスティックバイアス課題(15問)それぞれの正答率

- 課題の難易度のばらつきが激しい
- バイアスに影響を受けやすいことを示す他の研究と大幅に異なる結果

3.2. Table2 : 本研究での主な変数間の相関関係(全て $p < .01$)

- 合理的思考の 2 つの構成要素(バイアスの回避、信念から独立した三段論法推論)はお互いに緩やかな相関を示す→合理的思考得点

- CRT は、合理的思考得点、ヒューリスティックバイアス、CA2 と高い相関がある
 - CRT が認知能力と緩やかに重なること、それが合理的思考の予測因子であることを明らかにした
- 合理的思考得点において、最も相関が高いのが CRT、次が認知能力
- 実行機能尺度と思考傾向尺度は小さいが有意な相関を示した
- ヒューリスティックバイアス課題と三段論法推論課題は予測因子のパターンが似ている
 - 両者とも CRT と強い相関がある
- ほとんどの尺度は、三段論法推論課題のパフォーマンスよりも、ヒューリスティックバイアス課題のパフォーマンスとより相関がある

3.3. Table3：回帰分析

- どのように CRT の予測分散が認知能力や実行機能尺度、思考傾向の予測分散と重なるのか

3.3.1. 基準変数：合理的思考合成得点

- 知能テスト、実行機能尺度、思考傾向尺度、CRT の順にブロックに入れた
 - ◇ 全て有意(知能テスト・実行機能尺度・CRT ; $p < .001$ 、思考傾向尺度($p < .05$))
- 合理的思考を予測する CRT 能力は、認知能力と共通した分散にも実行機能と共通した分散にも起因しない
- CRT に相当量の固有分散(unique variance)
 - ◇ 次に大きい知能テストの固有分散の 2 倍近くある

3.3.2. 基準変数：ヒューリスティックバイアス課題得点

- 合理的思考合成得点の分析と同様の順で入れた
 - ◇ 全て有意($p < .001$)
- ヒューリスティックバイアスを予測する CRT 能力は、認知能力や実行機能、思考傾向と共通した分散に起因しない
- CRT は最も大きい固有分散を予測する変数だが、この分析で次に有力な固有分散は思考傾向だった

3.3.3. 基準変数：三段論法推論得点

- 合理的思考合成得点の分析と同様の順で入れた
 - ◇ 知能テストと CRT のみ有意($p < .001$)
- CRT は最も大きい固有分散を予測する変数だが、この分析では知能テストも同等の固有分散を予測した

※ 上記の分析では実行機能尺度は強い固有分散がないが、先行研究では認知能力とワーキングメモリに強い関連を示している

- Table2 では、知能テスト単体よりもワーキングメモリを含んだ認知能力尺度(CA2)は合理的思考パフォーマンスとより高い相関を示している

3.3.4. 基準変数：合理的思考合成得点

- 知能テストとワーキングメモリ得点、思考傾向尺度、CRT の順に入れた
 - ◇ 全て有意(知能テストとワーキングメモリ得点、CRT ; $p < .001$ 、思考傾向尺度 ; $p < .05$)
- CRT は 3 つの中で最も有力な固有予測因子

3.4. Table4：共通性分析

- 合理的思考の予測因子としての変数の部分的重複を明らかにするため
- 変数：合理的思考パフォーマンスを説明する3つの分散(CRT、拡大した認知能力、思考傾向)
- 3グループの予測因子による分散の合計：35.6%

4. Discussion

- CRTは認知能力と合理的思考能力両方と緩やかに関連している
 - 合理的思考変数との相関はいくつもの予測因子の間で最も高い
 - CRTは認知能力と重複しているので、認知能力を通して予測力を蓄えることも可能
- 回帰分析は、CRTが知能だけでなく実行機能と思考傾向から独立している合理的思考パフォーマンスを予測できたことを示した
 - CRTは知能尺度より多くの固有分散を占めた(Table3)
- CRTは基準変数の分散を実行機能尺度より一貫して予測した
 - CRTで良い成績を出すことは、最近の実行機能の研究で主張されるセットシフトや抑制統制特性を強調するはず
 - 原因：今回の実行機能尺度はとても不明確で異質だった可能性
 - もし抑制に焦点を置いたり複数の課題を使用したりしたなら、実行機能とCRTに重複が発見できたかもしれない
- CRTは相当量の信頼性のある分散を説明することが出来る
 - なぜそうなったのか→合理的思考の誤りについての分類体系の観点から考える (Stanovich, Toplak, & West, 2008)
 - ◇ 人の脳は二つの広い特徴がある(問題処理と問題内容)という発見が基盤
 - ◇ そして知能は両方に反して不適切な感化を与える
- 問題処理について：人は認知的儉約家(cognitive misers)の傾向にある
 - 判断と意思決定の認知科学における主な研究テーマ
 - なぜ認知的儉約家になるのか：人はもともと認知的負担を軽くするヒューリスティック処理構造の傾向にある
 - ヒューリスティック処理はしばしば適切な反応を早く与えるが、近代はそれよりも正確な思考を必要とする
 - 認知的儉約家になることは人々を目的達成から遅らせる
 - ◇ ヒューリスティックバイアスの文献におけるいくつかの効果は、人の儉約的処理の傾向の結果を示す
- 問題内容について：人はより合理的でない
 - 思考プロセスを覆すことは、知識や方略を用いて内容を役立たせる
 - 合理的思考の誤りは知識のギャップが原因で、潜在的にある論理的な知識基盤において起こる
- ヒューリスティックバイアス課題のパフォーマンスの予測因子としてのCRTの可能性は、知識のギャップを評価する能力から派生していない
- 一方で、認知的儉約家な人の思考とCRTは高い関連があるように思える

- CRT において、誤った回答は最初に用意されるが、倏約的処理は誤った回答を覆さず、上位の反応によって取り替える
- CRT は、合理的思考を支える基礎能力や予測因子というより、部分的な合理的思考尺度
 - CRT は倏約的処理から起こる推論エラー傾向を測定する尺度
 - 論理的な構造のため、倏約的な傾向を測定する非常に有力な尺度かもしれない
 - ◇ 間接的に測定する主観評定尺度ではなく、直接倏約的処理を測定できるパフォーマンス尺度

Table 1 Percentages of correct responses on each of the heuristics-and-biases tasks

Task	% Correct Responses
1. Causal Base Rate	53.8
2. Sample Size: Hospital Problem	28.6
3. Sample Size: Squash Problem	15.6
4. Regression to the Mean	40.2
5. Gambler's Fallacy 1	69.4
6. Gambler's Fallacy 2	92.2
7. Conjunction Problem	19.1
8. Covariation Detection	36.1
9. Methodological Reasoning	27.2
10. Bayesian Reasoning	22.3
11. Framing Problem	62.7
12. Probabilistic Reasoning: Denominator Neglect	61.8
13. Probability Matching	23.5
14. Sunk Cost	64.2
15. Outcome Bias	72.0

Table 2 Correlations between Cognitive Reflection Test, rational-thinking tasks, cognitive ability measures, executive-function measures, and thinking dispositions measures

Task	1	2	3	4	5	6	7	8	9	10	11	12	13
1. Cognitive Reflection Test	–												
Rational-Thinking Measures													
2. Heuristic-and-Biases Composite	.42	–											
3. Belief Bias Syllogistic Reasoning	.36	.29	–										
4. Rational-Thinking Composite	.49	(.80)	(.80)	–									
Cognitive Ability Measures													
5. WASI Composite	.32	.32	.35	.41	–								
6. WASI Vocabulary	.17	.24	.32	.35	(.76)	–							
7. WASI Matrix Reasoning	.32	.24	.21	.28	(.76)	.16	–						
8. CA2 (WASI Comp+ WM)	.40	.39	.36	.47	(.81)	(.66)	(.71)	–					
Executive-Function Measures													
9. Set Shifting	.28	.15	.12	.17	.27	.11	.30	.40	–				
10. Inhibition	.15	.17	.16	.21	.24	.18	.19	.35	.32	–			
11. Working Memory	.33	.33	.22	.34	.30	.17	.29	(.80)	.41	.36	–		
Thinking Dispositions													
12. Actively Open-Minded Thinking	.10	.18	.12	.19	.21	.25	.07	.17	.07	.09	.02	–	
13. Superstitious Thinking	.04	.24	.04	.18	.12	.22	.04	.15	.08	.10	.14	.38	–
14. CFC	.10	.16	.15	.19	.26	.23	.16	.21	.06	.04	.03	.21	.12

Signs have been changed so that all correlations indicate values in expected direction. Correlations in parentheses reflect part-whole relationships. $r = .088$ is significant at $p < .05$; $r = .125$ is significant at $p < .01$; $r = .165$ is significant at $p < .001$, one-tailed

Table 3 Regression results

	Hierarchical		Unique Variance Explained
	R ² Change	F to Enter	
Criterion Variable = Rational-Thinking Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix	.173	35.98	.042
2. Executive-Function Block: Set-S; Inhibition; WM	.056	8.26	.025
3. Thinking Dispositions Block: AOT; Super; CFC	.021	3.12	.017
4. Cognitive Reflection Test	.112	59.01	.112
Overall Regression: $F = 21.22$			
Multiple $R = .602$			
Multiple $R^2 = .362$			
Criterion Variable = Heuristics-and-Biases Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix	.100	19.09	.013
2. Executive-Function Block: Set-S; Inhibition; WM	.060	8.07	.026
3. Thinking Dispositions Block: AOT; Super; CFC	.042	5.96	.040
4. Cognitive Reflection Test	.080	37.45	.080
Overall Regression: $F = 14.68$			
Multiple $R = .531$			
Multiple $R^2 = .282$			
Criterion Variable = Belief Bias Syllogistic Reasoning			
1. Intelligence Block: WASI Vocab; WASI Matrix	.128	25.10	.051
2. Executive-Function Block: Set-S; Inhibition; WM			
Set-S; Inhibition; WM	.019	2.49 ^{ns}	.009
3. Thinking Dispositions Block: AOT; Super; CFC	.008	1.03 ^{ns}	.005
4. Cognitive Reflection Test	.064	27.68	.064
Overall Regression: $F = 10.44$			
Multiple $R = .468$			
Multiple $R^2 = .219$			
Criterion Variable = Rational-Thinking Composite			
1. Intelligence Block: WASI Vocab; WASI Matrix; WM	.227	33.52	.074
2. Thinking Dispositions Block: AOT; Super; CFC	.021	3.14	.016
3. Cognitive Reflection Test	.108	56.71	.108
Overall Regression: $F = 26.70$			
Multiple $R = .597$			
Multiple $R^2 = .356$			

WASI = Wechsler Abbreviated Scale of Intelligence; Set-S = set shifting; WM = working memory test; AOT = Actively Open-Minded Thinking scale; Super = superstitious thinking; CFC = Consideration of Future Consequences scale. ^{ns} Nonsignificant F value; all other F values reached significance

Table 4 Results of a commonality analysis using the rational-thinking composite score as a criterion variable

	1. CRT	2. Cognitive Ability: Vocab; Matrix; WM	3. Thinking Dispositions: AOT; Super; CFC
Unique Variance	.108	.074	.016
Common 1 & 2	.102	.102	
Common 1 & 3	.005		.005
Common 2 & 3		.028	.028
Common 1 & 2 & 3	.023	.023	.023
Total Unique + Common Variance	.237	.227	.071

CRT = Cognitive Reflection Test; Vocab = Wechsler Abbreviated Scale of Intelligence (WASI) Vocabulary subscale; Matrix = WASI Matrix Reasoning subscale; WM = working memory test; AOT = Actively Open-Minded Thinking scale; Super = superstitious thinking; CFC = Consideration of Future Consequences scale; Unique Variance = explained variance in rational-thinking composite scores that is unique to that variable block; Common 1 & 2 = explained variance that is common to Variable Blocks 1 and 2; Common 1 & 3 = explained variance that is common to Variable Blocks 1 and 3; Common 2 & 3 = explained variance that is common to Variable Blocks 2 and 3; Common 1 & 2 & 3 = explained variance that is common to all three variable blocks.