

Communicating and Distorting Risks with Graphs: An Eye - Tracking Study.

Margo M. Woller – Carter, Yasmina Okan, Edward T. Cokely, Rocio Garcia – Retamero, 2012, Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting, 1723-1727.

Introduction

- 私たちは複雑で不確かな世界を理解するためのリスクに情報を用いる
 - 先行研究では、多くの人が情報を正確に解釈し行動することを失敗すると述べている
- 最近の先行研究は、異なる能力と異なるリスク提示形式が、リスク理解の影響をどのように整えるのかに焦点をおいている
- さらに、テクノロジーの進歩がリスクについての情報の利用可能性を増やすおかげで、リスクコミュニケーションにとってより効率的でシンプルで分かりやすいツールを発展させる結果が広まりつつある

- 図やグラフが複雑な数的情報のコミュニケーションを促進することができるという信念は広く受け入れられている
 - 画像表示のようなツール（線グラフや散布図などを含む）は専門家と社会によって作られたエラーを弱める手助けをする
- 1998年、HFES（the Human Factors and Ergonomics Society）は定量的データを示すグラフガイドラインを発行した
 - グラフ利用とデザインの数少ないガイドラインの1つ
 - ◇ x軸が時間を示すとき、情報はいつも左から右へ年代順に示される
 - ◇ x軸の数的情報は低いところから高いところへ示される
 - しかし、デザイナーはしばしば、グラフをデザインするとき、それらのガイドラインに従わない

- 薬剤広告グラフのデザイン特徴の調査（Cooper, Schringer, Wallace, Mikulich and Wilker, 2003）
 - 74個のグラフサンプルの内、40%が歪められていた
 - ◇ 3Dの乱用（20%）、不適切なスケールや分割された軸（16%）、不適切な基線（12%）
 - 不適切なデザインのグラフは、解釈のエラーを導き、次に、判断に影響する

- 先行研究は、人がバイアスにより情報を理解する能力に違いがあることを明らかにした
 - グラフリテラシースキルが高く、数的スキルが相対的に低い個人
 - ◇ →数字だけより数字とグラフで表された方がより良いパフォーマンスをする
 - グラフリテラシースキルも数的スキルも低い個人
 - ◇ →数字だけの場合か、数字とグラフ両方で表された場合か、どちらか一方で情報を理解することに苦勞する

- 本論文では、デザインの乏しいグラフの理解に影響する認知メカニズムを調査した
- 政治と医学、2つのリスクコミュニケーションのグラフを使用
 - 1つ目は、オバマ政権の経済政策を議論するニュースからのグラフ
 - 2つ目は、インフルエンザの蔓延を抑える抗ウイルス剤の広告からのグラフ
 - 両グラフとも HFES デザインガイドラインに違反している
- まず、グラフが解釈的エラーと関連しているかどうかを評価
 - 次に、エラーといくつかの認知能力（ニューメラシー、グラフリテラシー、認知的内省）の関係を検討

Experiment 1A

- 先行研究は、グラフリテラシーとニューメラシーの個人差がリスク理解の差を予測することを示す
- 実験 1A の仮説
 - H1：2つのグラフは先行研究通り判断エラーを引き起こす
 - H2：認知能力の個人差は、パフォーマンス（エラー率）の差を予測する
 - H3：パフォーマンスの個人差は、刺激の記憶の差に関連している
 - H4：記憶の差は、少なくとも認知能力とエラー率の間の関係を部分的にとりなしている
- もし認知能力が刺激の精巧なエンコードを予測することで、グラフ解釈を促進するならば、より高いパフォーマンスをする個人は、より刺激を記憶している

Method

- Participants
 - 心理学入門クラスの大学生 52 名
 - ◇ $M_{age}=21$ 、 $SD_{age}=2.8$ ：男性 34 名、女性 16 名、性別無記入 2 名
- Materials
 1. 政治的グラフと医学的グラフについての判断問題
 2. それぞれのグラフについて抜き打ち記憶テスト
 - ◇ 政治的領域のグラフ (Figure1)
 - ✓ x 軸が時系列順になっていない
 - 3つの判断問題と 12 個の記憶テスト
 - x 軸の変数はどんな時間単位か？ - 選択肢：週、月、年
 - ◇ 医学的領域のグラフ (Figure2)
 - ✓ 読み手に誤解を与える突出した情報を提供している
 - 4つの判断問題と 13 個の記憶テスト
 - y 軸の値はどんな間隔で増加するか？ - 選択肢：3,5,10
 3. 認知内省課題
 - ◇ 認知的熟慮性と衝動性の尺度 (Frederick,2005)

4. ベルリンニューメラシーテスト
 - ◇ 統計的ニューメラシーとリスクリテラシーの適応尺度 (Cokely et al., 2012)
 5. グラフリテラシー尺度
 - ◇ グラフ理解（データを読み取ること、ポイントを比較すること、傾向を読み取り理解すること）の尺度 (Galesic & Garcia-Retamero, 2011)
- Procedure
 1. アイトラッカーのキャリブレーション
 - ◇ ただし全ての被験者からデータを取ったわけではない（詳しくは実験 1B で）
 2. グラフ提示
 - ◇ グラフ提示順序は被験者間でカウンターバランスを取る
 3. 判断問題
 4. 抜き打ちの記憶テスト
 - ◇ 記憶問題項目の提示順序（政治的関連か、医学的関連か）はグラフ提示の順序に従う
 5. 尺度とアンケート
 - 実験はおよそ 15 分

Results & Discussion

- パフォーマンス：情報を正確に解釈していることと定義
 - 記憶得点：記憶テストの正解総数
- Performance
 - H1：chance performance と実際の判断を比較する
 - ◇ 政治的、医学的両課題のパフォーマンスは約 9.6%、chance performance より高いが、有意ではない ($t(51)=1.40, p=.17$)
 - ◇ この結果は、参加者がグラフを誤って解釈する傾向にあることを示す
 - Cognitive abilities and Performance
 - H2：片側二変量相関 (Table1)
 - ◇ 政治的課題のパフォーマンスは、ニューメラシー、グラフリテラシー、記憶と相関がある
 - 認知的内省は、政治的課題パフォーマンスと有意ではない
 - ◇ 医学的課題のパフォーマンスは、全てにおいて相関がある
 - Memory and Individual differences
 - H4：パフォーマンスと個人差間の関係を実験し、階層的線型回帰モデルを構築
 - ニューメラシー、グラフリテラシー、認知的内省をモデル 1 に、記憶をモデル 2 にして分析
 - ◇ 政治的課題：フルモデルはパフォーマンスの強い予測因子 ($F(4,47)=4.33, p=.005, R^2=.27$)
 - 予測因子のラインにおいて、記憶は固有分散の多くの割合を占める ($t=2.27,$

$$p=.03, \beta=.37)$$

- 記憶に起因するとき、個人差とパフォーマンスの関係は有意ではなくなる
- ◇ 医学的課題：フルモデルはパフォーマンスの強い予測因子 ($F(4,47)=4.22, p=.005, R^2=.26$)
 - 予測因子のラインにおいて、記憶は固有分散の多くの割合を占める ($t=2.47, p=.02, \beta=.38$)
 - 記憶は、ニューメラシー ($t=.71, p=.48, \beta=.10$)、グラフィテラシー ($t=.95, p=.35, \beta=.13$)、認知的内省 ($t=.17, p=.87, \beta=.03$) とパフォーマンスの関係を十分とりなした
- 認知能力が乏しいデザインのグラフのパフォーマンスを予測する理論と一致

Experiment 1B

- 先行研究では、タスクパフォーマンスの違いは、記憶の個人差による
 - しかし、記憶は、グラフをより早く識別し、思い出すスキルと単純に関連しているだけかもしれない
- 専門知識は記憶のエンコードと検索を高める
 - チェスマスターは駒を動かす際に選択肢を探索する必要がない
 - ◇ より複雑な認知的表象を持ち、より効率的な情報を記憶しているため
- 先行研究で観察された記憶の違いが、より精巧な情報探索に関係するかどうかを評価するために、実験 1B では実験中の参加者の視線と反応時間を分析
- 実験 1B の仮説
 - H5：記憶は参加者の視線によって測られる精巧な情報探索の差と強く関連する (Cokely & Kelley, 2009)

Method

- Participants
 - 実験 1A を終えた大学生 21 名 ($M_{age}=21, SD_{age}=2.5$: 男性 13 名、女性 8 名)
- Materials
 - faceLAB 4.6 Seeing Machines、Gaze Tracker 8.0 を使用
 - 注視点区切りは 250ms
- Procedure
 - 実験 1A とほぼ同様の手続き
 - キャリブレーション後に実験を行った
 - ◇ 政治的グラフのパフォーマンスが、視線計測なしの場合とありの場合で異なった ($F(1,50)=7.35, p=.009$)
 - ◇ 対照的に、医学的課題と両グラフの記憶テストのパフォーマンスは、視線計測なしの場合でもありの場合でも異ならなかった

Results & Discussion

- 両グラフの AOIs (areas of interest) を調べる
 - データ、x 軸、y 軸の 3 領域
 - ◇ 政治的グラフ：
 - 正確な解釈に至るために必要な情報→ x 軸
 - ◇ 医学的グラフ：
 - 正確な解釈に至るために必要な情報→ y 軸とデータ
- 情報探索は、両グラフとも以下の点で評価される
 1. AOI における注視点の総数
 2. AOI における総時間
 3. グラフ全体における注視点の総数
 4. グラフ全体において費やした総時間
- Memory and search
 - H5：片側二変量相関
 - 政治的課題の記憶パフォーマンス
 - ◇ x 軸における注視点 ($r(19) = .37, p = .05$) と時間 ($r(19) = .39, p = .04$) の両方に有意な相関があった
 - 医学的課題の記憶パフォーマンス
 - ◇ y 軸の注視点 ($r(19) = .40, p = .04$) と時間 ($r(19) = .40, p = .04$)、データの注視点 ($r(17) = .40, p = .05$) と時間 ($r(17) = .40, p = .05$)、グラフ全体における注視点 ($r(19) = .51, p = .009$) と時間 ($r(19) = .49, p = .01$) に有意に相関があった
- Search and performance
 - 片側二変量相関
 - ◇ 媒介分析は、サンプルが少なく確実でないので行わない
 - 政治的課題パフォーマンスの分析：
 - ◇ 政治的グラフの y 軸の注視点とはわずかに有意 ($r(17) = -.37, p = .06$) で効果量も大きかった ($d = .93$) が、記憶や情報探索との有意な相関は検出されなかった (Figure3)
 - 医学的課題パフォーマンスの分析：
 - ◇ 記憶 ($r(19) = .46, p = .02$) とグラフの注視点の総数 ($r(19) = .40, p = .04$) と有意に関連がある
 - ◇ 情報探索は精巧なエンコードとパフォーマンス間の関係を取りなした

General Discussion

- リスクコミュニケーションと数的情報（特に産業技術領域、政治領域、医学領域、金融領域）の透明化のニーズが高まっている
- グラフは効果的なコミュニケーションの一つの良い手段を示す
 - しかし、今回の結果から、乏しいデザインのグラフは、判断と理解の主なエラーに関

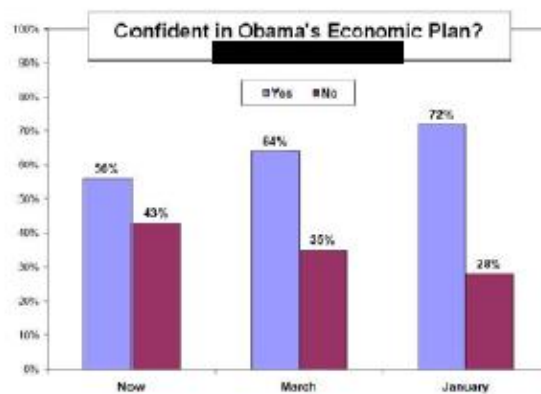
連する傾向があることが示される

- 結果はさらに、ニューメラシーやグラフィテラシー、認知的内省のようないくつかの個人差がエラーを防ぐことを示した
- 同様に、パフォーマンスにおける認知能力の影響は、記憶の差によって仲介され、視線計測によって測られた精巧な情報探索と関連する

Implications & Future Directions

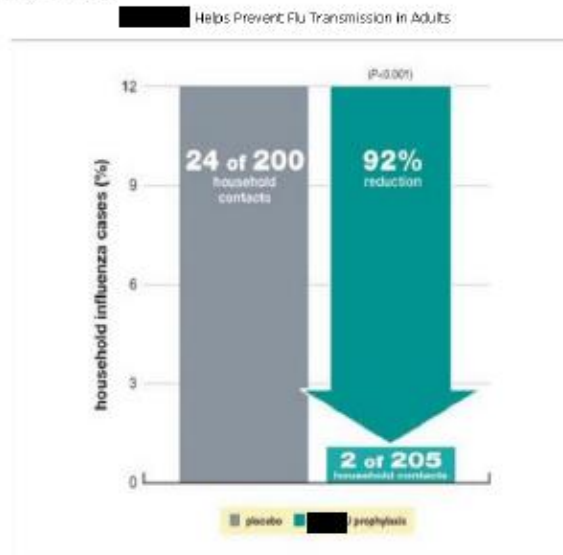
- 今回の結果は、情報探索、認知能力、グラフ理解の間のリンクにおける初めてのアイトラッキングのいくつかの証拠を示す
- 結果は、先行研究の発見（グラフィテラシーが、人がタイトルやラベルの必要な情報やグラフの解釈のスケールを観点に含む可能性に影響すること）にまとめられた
- 今回は限界のサンプル量と類似の実験デザインだったため、今後の研究では、情報探索の違いが実際に記憶や理解を促進するのかを調査する必要がある
 - 結果は2つの強いつながりを示したけれども、精巧なエンコードが理解の促進に必要もしくは十分かどうかははっきりしていない
- グラフデザインの優先目標は分かりやすいグラフを作ること
- HFESのガイドラインはグラフ発展の役に立つ基盤を提供する
 - 人のプロセスの限界とデザインの本質を組み込んでいる
- 今後のデザインガイドラインはユーザーに合わせたグラフを作るという目標が含まれる
 - 人的要因、認知科学、意思決定科学、ユーザビリティの共通の働きが必要
- 残念ながら、今回の結果によるハイライトによれば、良いガイドラインが整っていても、使われないかもしれない
- グラフを歪めるいくつかの理由
 - ガイドラインの認識の欠如
 - 潜在的な動機のような関心の不一致
 - ◇ 医学的判断課題にて、代替薬品の価格と有効性を考慮してグラフに記載された薬品の価値を評価させた
 - ◇ →参加者のほとんどがグラフに示された薬品を過大評価した
 - 端的に言えば、代替薬品と比較し、2倍の価値を判断した
- 関心の不一致と規則の限界ゆえに、グラフは消費者やデザイナーに力や情報を与える簡単なツールの必要がある
- 今回の調査は、インターネットに提示されたグラフの質を評価することで用いるオンライン適応テストを発展している
- このインターネットをベースとしたツールは、簡単な質問を訊ね、例を提示し、グラフの質の推測を与える

- 特に、ソフトウェアはグラフの質の関係についてのフィードバックを与え、最小限必要なグラフリテラシーのレベルを記載する
- もし、正確で信頼できるフィードバックを与えることに成功したら、それはグラフガイドラインのユーザビリティやグラフデザイナーのトレーニングを促進することに使われるだろう
- グラフリテラシーの重要さの認識と良いグラフデザインのメリットを促進できることを期待する



According to the graph...
 1) Has confidence in Obama's Economic Plan increased?
 Yes No

Figure 1. Redacted political graph and the primary performance question.



According to the graph...
 prevents flu infections for 92 out of 100 people (92%).
 True False

Figure 2. Redacted medical graph and the primary performance question.

Table 1. Correlations for cognitive abilities and performance.

	Political Performance	Medical Performance
Cognitive Reflection	.07	.26*
Numeracy	.29*	.29*
Graph Literacy	.36**	.33**
Memory	.47**	.49**

*Note: * $p < .05$, ** $p < .01$.*

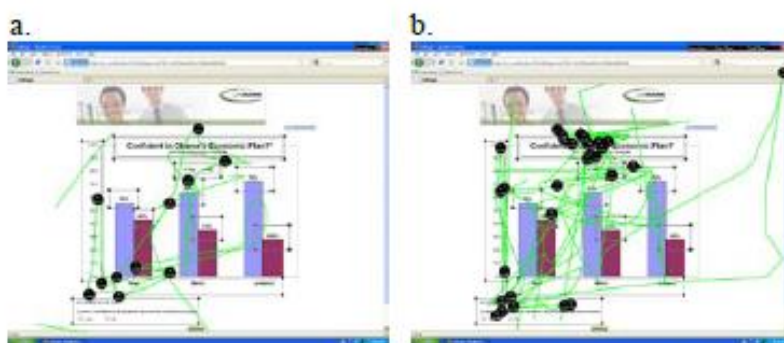


Figure 3. Example scanpaths for the political graph; (a) low information search scanpath and (b) high information search scanpath.