

文書の意見と信憑性がユーザの検索行動および 信念の変化に与える影響の分析

ポチラッタナチャイクル スパナット[†] 山本 岳洋[†] 山本 祐輔^{††} 吉川 正俊[†]

[†] 京都大学 大学院情報学研究科 〒606-8501 京都府京都市左京区吉田本町

^{††} 静岡大学 大学院総合科学技術研究科 〒432-8011 静岡県浜松市中区城北 3-5-1

E-mail: [†]{suppanut, tyamamot}@dl.kuis.kyoto-u.ac.jp, yoshikawa@i.kyoto-u.ac.jp,

^{††}tyamamoto@inf.shizuoka.ac.jp

あらまし 本研究では、Web 検索中に検索ユーザが閲覧する文書の意見と信憑性という2つの要因に着目し、これらの要因が検索時の行動や検索後の最終的な意思決定に与える影響を分析した。具体的には、「乳製品は高血圧の改善に効果的か」といった、yes-no 型での解答が求められる検索タスクを被験者に与え、被験者に返す検索結果を人工的に操作することで、これらの要因が与える影響を分析した。275名の被験者から得られたデータを分析した結果、被験者は、検索前にいっていた検索タスクに対する yes か no かという考え（信念）と相反する意見を含んだ文書を閲覧すると、検索プロセスにより労力をかけ、検索後に検索前とは反対側の信念へと変える可能性が高くなることが分かった。本研究で得られた知見は、ユーザの事前の信念と相反する文書を検索結果の上位に配置することで、ユーザが慎重に情報を収集する可能性があることを示唆している。

キーワード 情報検索, Web 検索, 信憑性, 確証バイアス

1 はじめに

Web 検索時に正確な情報を集め意思決定することは重要ではあるが難しい。正確な情報を獲得するためには、多様な情報に触れること [9] [13] や、できる限り検索に時間を書けて情報を獲得したり、文書の書き手が誰であるかを意識しながら閲覧したりしながら、検索時に慎重に情報を精査すること [17] [19] が重要であると言われている。しかし、別の研究によれば、検索時にユーザは自らの信念 (belief) に沿った情報のみを探すという、認知バイアスの一種である確証バイアスの影響を受けてしまうことが知られている [15]。確証バイアスに強く影響を受けてしまうと、前述した慎重に情報を精査するような行動をとることは難しい。システムがユーザの興味に合わせて情報をランキングしたりフィルタリングすることでユーザが閲覧する情報が偏ってしまう状況は近年ではフィルターバブルと言い、その危険性が指摘されている [11]。

一方で、ユーザが獲得する情報の意見や質がユーザの確証バイアスや情報行動に影響を与えることを示す実験結果が報告されている [4] [6] [10]。たとえば、Schwind らはユーザの興味と一致しない情報をそのユーザに提示することで、確証バイアスを抑制することができると報告している [14]。こうした研究は、Web 検索時においても、ユーザが情報をより精査しながら検索するようなランキングが実現可能であることを示唆していると考えられる。

そこで本研究では、検索結果中にユーザが閲覧する文書の意見やその質に着目し、これらの要因がユーザの検索行動や意思決定に与える影響を分析する。具体的には、「乳製品は高血圧の

改善に効果的か」といった yes-no 型での解答が求められる検索タスクに注目し、ユーザが検索前に検索タスクに対していどく、乳製品は高血圧に効果的であると信じるかどうかという事前の信念と文書の内容が一致するか相反するかという文書の意見 (opinion)、また、その文書がどの程度信じられると感じるかという文書の信憑性 (credibility) という2つの要因に着目する。本研究の目的は、これらの要因とユーザの検索行動や意思決定との関係を明らかにすることで、ユーザがこれまでよりも慎重に情報を精査するようなランキングをどのように実現すべきかという問いに対する示唆を得ることである。

本研究で取り組む研究課題は以下のとおりである：

RQ1: 検索時に閲覧する文書の意見と信憑性が、その検索における検索行動にどのような影響を与えるか。

RQ2: 文書の意見と信憑性が、検索前後の信念の変化にどのように影響を与えるか。

RQ3: 信念の変化の有無と検索行動に関連はあるか。

これらの研究課題を明らかにするため、クラウドソーシングによるユーザ実験を行った。実験ではまず被験者に検索タスクに対する事前の信念をアンケートにより回答してもらう。その後、被験者の事前の信念に合わせて、我々が用意した文書を被験者が閲覧する検索結果中に人工的に挿入する。そして、検索タスク終了後に再度ユーザに対して検索タスクに対する信念を獲得する。これにより、ユーザが検索中に遭遇する文書の意見や信憑性をコントロールでき、異なる意見や信憑性がユーザの検索行動や信念の変化に与える影響を分析することができる。

275名の被験者から得られたデータを分析した結果、主に以下のような知見を得た。(1) 被験者は検索前の信念と相反する

意見を含んだ文書に遭遇すると、より多くのクエリを投入し、多くの文書を閲覧し、より多くの文書を最終的な意思決定の証拠としてあげるなど、検索プロセスにより労力をかけた。(2) また、検索前の信念と相反する意見を含んだ文書に遭遇すると、検索後に検索前とは反対の信念へと変える可能性が高くなった。(3) さらに、この信念の変化と(1)で見られたような検索行動の違いは認められなかった。これらの知見は、ユーザの事前の信念と相反する文書を検索結果の上位に配置することで、ユーザが慎重に情報を収集する可能性があることを示唆しているとともに、ユーザの興味や嗜好に基づくランキングがユーザの確証バイアスをより強くする可能性があることを示唆している。

2 関連研究

これまでに、ユーザの検索行動や意思決定が検索結果中のバイアスや、文書の信憑性に影響を受けることを示唆する研究が報告されている。White は検索プロセスを通じてユーザの確証バイアスが強まっていくこと [15] [16]、Pogacar らは検索結果の多くが不正確な情報を含んでいた場合、検索ユーザの検索後の信念も不正確なものになってしまう可能性が高くなることを明らかにしている [12]。また、Kammerer らは [7] ユーザは検索結果の信憑性を高いと判断した場合、あまり労力をかけずに検索結果の評価を行うことを報告している。さらに、1 章で述べたように、情報の意見や質がユーザの確証バイアスに与えることを調査した研究もなされてきた [4] [6] [10] [14]。

一方で、検索中に閲覧する文書の意見や質（信憑性）が、クエリ投入やクリック、文書閲覧といった詳細な検索行動にどのような影響を与えるのか、また、それらの行動と意思決定との関連などについては明らかにされていない。本研究では、こうした点を明らかにすることで、検索プロセス中にユーザがより慎重に情報を精査する要因についてより理解するための知見を得る。

3 実験デザイン

本章ではまず、検索時に正しい情報を獲得するために必要だと考えられる検索行動について説明する。つぎに、本研究で検証する文書の意見と信憑性について詳しく述べる。最後に、本研究で我々が実験のために用意した、それぞれの意見と信憑性に応じた文書をどのように作成したのかについて説明する。

3.1 批判的検索行動

Ennis は批判的思考を、何を信じどのように行動するかという意思決定における慎重な思考プロセスと定義している。批判的思考に基づいて行動するユーザは、代替案を探し、さまざまな観点から思考し、そして自らの信念やバイアスについて自覚することが必要であると述べられている [3]。Yamamoto らはこの Ennis の批判的思考の定義に基づいて、Web 検索において批判的思考を発揮し情報を精査するユーザに求められる行動を議論している [17] [18] [19]。本研究では、そのような行動を批判的検索行動と呼ぶ。批判的検索行動の中には、より検索セッ

ションに時間をかけて情報を集める、さまざまなクエリを投入して多様な情報を集める、書き手が誰であるかを意識して文書を閲覧する、最終的な意思決定をサポートするためにより多くの文書を証拠として集める、といったものが含まれている。具体的には、本研究では以下の行動を批判的検索行動として捉える。

- より多くのクエリの投入して情報を収集する。
- より多くの文書を閲覧する。
- より下位にランキングされた文書まで閲覧する。
- より多くの時間をかけて文書を閲覧する。
- 検索タスクにより多くの時間をかける。
- より多くの情報を意思決定の判断に考慮する。

本研究の目的は文書の意見や信憑性といった要因がこれらの行動にどのような影響を与えるのかを明らかにする (RQ1) とともに、信念の変化 (RQ2) とこれらの行動との関係を分析する (RQ3) ことである。

3.2 検証する要因

本研究では、文書の意見と信憑性という 2 つの要因に着目する。まず、文書の意見について説明する。本研究では、文書の意見を、検索ユーザの事前の信念に対する極性と定義し、ユーザの事前の信念と一致 (**consisitent**) する意見、相反 (**inconsistent**) する意見という 2 種類の意見を考慮する。たとえば、「乳製品は高血圧の改善に効果的か」という検索タスクに対してユーザが検索タスク開始前に、乳製品は高血圧の改善に効果的であると信じているとき、文書の内容が「乳製品は高血圧に効果的であると示す根拠はない」ことを主張しているとき、その文書の意見はユーザの事前の信念と相反した意見である。

つぎに、文書の信憑性について説明する。本研究では、Fogg らの定義 [5] に従い、信憑性 (credibility) をその情報や情報源が信じられる度合い (believability) とする。具体的には、文書の信憑性として高信憑性 (**high credibility**) と低信憑性 (**low credibiltiy**) の 2 種類を考える。

3.3 制御文書の作成

本研究の目的は、前節で述べた要因が検索行動や信念の変化に与える影響を分析することである。この目的を達成するため、本研究ではそれぞれの要因に応じた Web 文書を作成し、その文書をシステムがユーザに返す検索結果ページ中に人為的に混ぜる。本研究ではそのような文書を制御文書 (**controlled documents**) と呼び、本節では制御文書の作成方法について述べる。4.1 節で述べるように、本研究の実験は被験者内計画にて行った。すなわち、文書の意見と信憑性それぞれの要因について 2 種類の水準を考えるため、 C_1 (相反, 高信憑性)、 C_2 (相反, 高信憑性, 低信憑性)、 C_3 (一致, 高信憑性) および C_4 (一致, 低信憑性) という 4 つの実験条件に対応する制御文書を各検索タスクに対して用意した。

まず、文書の意見について説明する。ユーザの意見と一致あるいは相反する意見を含む文書を用意するため、我々は検索タスクごとに、検索タスクについて支持する、すなわち「乳製品



(a) 高信憑性文書



(b) 低信憑性文書

図 1 実験で用いた制御文書のスクリーンショット。ある検索タスクの意見に関して、制御文書中の記事の内容は (a) 高信憑性、(b) 低信憑性ともに同一の記事を引用している。

は高血圧の改善に効果的である」と主張しているページと、支持しない、すなわち「乳製品は高血圧の改善に効果的であるとは言えない」と主張しているページを手で収集した。本研究では、支持、非支持どちらの意見にも信頼できる主張があるという前提のもと、以下の基準のどれかを満たすページを信頼できるページとして、ページを収集した。

- 科学技術論文もしくは学会報告を引用している記事
- 根拠に基づく医療のための医療ナレッジベース¹
- 人物もしくは団体が著者として明記されているページ

表 1 に実験で用いた検索タスクおよび収集したページの一覧を示す。

つぎに、文書の信憑性について説明する。信憑性の高い文書、低い文書を用意するため、本研究では表面型の信憑性 (surface credibility) [5] に着目し、高信憑性、低信憑性それぞれに応じたページの視覚的デザインを用意した。表面型の信憑性とは信憑性判断基準の 1 つであり、外見上の特徴によって情報の信憑性を判断するような判断基準である。たとえば、専門家がデザインしたように見える洗練されたデザインのページを信憑性が高いと判断することは、表面型の信憑性の例である。

図 1 に実際に用意したデザインを示す。高信憑性なページ

(図 1(a)) は整ったレイアウトでモダンなデザインである一方で、低信憑性なページ (図 1(b)) は素人がデザインしたようなレイアウトである。また、一般的に Web ページのドメインも情報の信憑性を判断するうえで重要な情報となるため、それぞれのデザインに合わせてドメインを用意した²。後で報告するように、被験者の多くは高信憑性なページの方がより信じられると回答しており (4.2 節参照)、今回用意したデザインによりユーザの信憑性判断をコントロールできたと考えられる。

具体的な制御文書の作成方法は以下のとおりである。検索タスクに対して支持する意見を含む高信憑性な制御文書は、表 1 に記載した、検索タスクを支持する情報源を引用する形で検索タスクに対して支持する主張をし、図 1(a) に示した視覚的デザインで表示する。一方、検索タスクを支持する低信憑性な制御文書は、同じ情報源を引用するが視覚的デザインを図 1(b) の形で提示する。ここで、高信憑性、低信憑性な文書両者とも、同一の情報限を引用するという形式で文書を作成しているため、視覚的デザイン以外は両者の内容はほぼ同一である。これは、高信憑性文書と低信憑性文書間で文書の主張に差異が生じてしまい、その差異が検索行動や意思決定に影響を及ぼしてしまうことを最小限に抑えるためである。本研究では 4 つの検索タスクを用いて実験を行ったため、合計 16 件 (4 検索タスク × 4 実験条件) の制御文書を作成し、次章で述べる実験に用いた。

4 実験

本章では実施したユーザ実験の詳細を述べる。まず実験計画について述べ、その後具体的な実施手順について説明する。最後に、被験者の募集方法および最終的に分析したデータについて述べる。

4.1 実験計画

本研究では 2 要因被験者内実験計画を用いた。すなわち、各被験者は意見 (一致・相反)、信憑性 (高信憑性・低信憑性) のすべての組み合わせによる 4 つの条件の下で実験する。被験者は、各条件に対して表 1 に示した 4 つの検索タスクから 1 つが割り当てられ、計 4 つの検索タスクを実施する。このとき、被験者が 4 つの検索タスクを行う順序については、実験条件およびタスクの両者を考慮した Graeco-Latin 方格 [8] に従い被験者を割り当て、実験条件およびタスクに関する順序効果を最小限に抑えた。

4.2 実験手順

まず、被験者に対して検索タスクの概要、収集するデータ、収集したデータの研究目的への利用などについて説明した。また、「タスクに関する説明に従っていないと判断した人については報酬を支払わない可能性がある」とあらかじめ説明を与えた。これらについて同意した被験者のみが実際の実験タスクへ進み、実験を行った。

2 : 高信憑性文書: <http://www.med-japan.com/>,

低信憑性文書: <http://www.kenko-blog-life.org/>

1 : <https://www.cochrane.org>

表 1 実験で用いた検索タスクおよび人工文書中で引用する情報源およびその情報源を選択した判断基準。

検索タスク		情報源	
ID	タスク文	支持	非支持
1	乳製品は高血圧の改善に効果的か。	http://bit.ly/2Tr2mYG (科学技術論文を引用)	http://bit.ly/2TmjsHo (科学技術論文を引用)
2	イソフラボンはコレステロール値の低下に効果的か。	https://nkb.jp/2CNTjM5 (学会報告を引用)	http://bit.ly/2Tshjdc (医療ナレッジベース)
3	高麗人参は認知症予防に効果的か。	http://bit.ly/2COPciQ (学会報告を引用)	http://bit.ly/2CPDV21 (医療ナレッジベース)
4	ホメオパシーはぜんそくの改善に効果的か。	http://bit.ly/2TrKGfA (実在の団体)	http://bit.ly/2Tq7Z9B (医療ナレッジベース)

つぎに、実際の検索タスクに入る前に、これから行う検索タスクへのイメージを深め使用する検索システムに慣れるために訓練タスクを行ってもらった。訓練タスクでは、「漢方薬はアトピー性皮膚炎に効果があるか」を後述する検索システムで調べてもらった。

訓練タスクの後、被験者に 4 つの検索タスクを実施してもらった。各検索タスクは以下の手順で実施した。まず、各検索タスクを開始するにあたり、検索タスクに関する背景およびタスクの内容に関する説明を被験者に提示した。たとえば、表 1 に示したタスク 1 に関しては、以下に示す説明文を被験者に提示した。

高血圧は心臓や血管系の病気になりやすく、また、高血圧な人はそうした病気になった際に死亡率が高いということが知られています。そのため高血圧を予防・改善することが健康のために重要であると考えられています。

あなたは現在高血圧の症状を持っており、医者に血圧を下げるよう言われています。そのような状況のなか、牛乳やヨーグルト、チーズなどの乳製品が高血圧の改善に効果があるらしいという噂を耳にしました。そこで、このタスクでは、ウェブ検索を用いて、乳製品が高血圧の改善に効果があるかどうかを自由に調べてください。

被験者に上記の説明文を読んでもらった後、タスク前質問に回答してもらった。タスク前質問では、検索タスクに対する事前の知識に加え、検索タスクに対して yes と信じるかか no と信じるかの信念の度合いを測定した。具体的には、「現時点で、乳製品が高血圧の改善に効果があるかと思いますか？」という質問項目に対して 4 段階のリッカート尺度 (1: そう思わない, 2: どちらかというと思わない, 3: どちらかというと思おう 4: そう思う) で回答してもらった。本研究では、この質問項目によって得られた、検索タスクに対する被験者の検索前の信念を**事前の信念** (prior belief) と定義する。

タスク前質問に回答してもらった後、被験者に実際に検索タスクを実行してもらった。このとき、被験者には次節で述べる検索システムだけを利用し、既存の検索エンジンは使用しないよう指示を与えた。検索タスク中は、被験者は自由に検索クエリをシステムに投入でき、また、検索結果をクリックすることで実際のページを自由に閲覧できた。また、検索タスクの制限時間は設けず、被験者が満足いく回答を得たと感じた時点で自由に検索タスクを終了してもらった。

被験者が検索タスクを終了した後、タスク後質問に回答してもらった。タスク後質問ではまず、検索タスクに対する信念を前述と同じ質問項目とリッカート尺度で回答してもらった。本研究ではこれを**事後の信念** (posterior belief) と定義する。つ



図 2 実験で用いた検索システムのスクリーンショット。図中の 2 位の検索結果は我々が用意した制御文書に対応する検索結果である。

ぎに、被験者の事後の信念の根拠となった文書を被験者に選択してもらった。具体的には、検索タスク中に被験者がクリックした文書の集合 (制御文書を含む) を被験者に提示し、被験者にその中から 1 つ以上の文書を選択してもらった。

4 つの検索タスクを終えた後、被験者に我々が用意した制御文書に対する信憑性を回答してもらった。具体的には、我々が用意した制御文書の高信憑性文書、低信憑性文書の両者を 4 つの検索タスクを通してクリックした被験者を対象に、高信憑性文書、低信憑性文書それぞれについて「このウェブページはどの程度信じられると思いましたか？」という質問項目に 4 段階のリッカート尺度 (1: 信じられない, 2: どちらかというと思えない, 3: どちらかというと思えられる, 4: 信じられる) で回答してもらった。なお、4.4 節で述べる分析対象に用いた 275 名の被験者による本質問項目に対する回答結果は、高信憑性ページに対しては平均 2.90 ($SD=0.80$)、低信憑性文書は平均 2.41 ($SD=0.75$) であり、Welch の t 検定にて統計的有意差が認められた ($p < 0.001, t = 9.31$)。

最後に、被験者のデモグラフィック属性に関する質問に回答してもらったと同時に、検索タスクに関して気づいた点を任意で自由に記述してもらった。これは、次節で詳しく述べるように実験で用いる検索システムは通常の検索結果をコントロールしているため、それに気づいた被験者の検索行動は今回の分析対象としては不適切と考え、分析結果から除くためである。

4.3 検索システム

図 2 に実験で用いた検索システムを示す。本検索システムは、画面上部に検索タスクに関する説明文、検索タスク開始からの経過時間、およびタスクを完了し次へ進むためのするためのボタンがある以外は一一般的な商用検索エンジンと同様のインタフェースである。

本システムは通常の検索結果に 3.3 節で述べた制御文書を混ぜた検索結果を被験者に提示する。被験者がいまクエリ q を入力したとき、システムはまず Bing Web Search API³を通じて q に対する Web 検索結果 50 件 $d_1, \dots, d_{50}$ を取得する。その後、現在の実験条件に合わせて、あらかじめ用意した制御文書 $d_{\text{controlled}}$ を検索結果の 2 位に挿入した 50 件の検索結果 $d_1, d_{\text{controlled}}, d_2, \dots, d_{49}$ を被験者に提示する。

挿入する制御文書 $d_{\text{controlled}}$ は、被験者が現在行っている検索タスクの実験条件とタスク前質問で測定した事前の信念に応じて決まる。たとえば、いま検索タスクが表 1 に示したタスク 1、実験条件が C_1 (相反, 高信憑性)、かつ被験者の事前の信念が「そう思わない」、あるいは「どちらかといえばそう思わない」であるとき、制御文書は事前の信念と相反した主張、つまり「乳製品は高血圧の改善に効果がある」と主張し、高信憑性文書用の視覚的デザインで表示される。なお、制御文書に対応する検索結果のタイトルに関しては文書の主張を端的に表すテキスト (例: 「乳製品は高血圧の改善に効果があるという研究報告」を我々が人手で用意し、それにウェブサイトのサイト名を加えたものを用いた。また、スニペットについては制御文書の記事中で引用しているテキストの先頭約 100 文字を用いた。

4.4 被験者

2018 年 7 月 10 日から 2018 年 8 月 3 日にかけて、クラウドソーシングプラットフォームであるランサーズ⁴を用いて被験者を募集し、401 名の被験者が実験に参加した。実験の開始から終了までにかかる時間は中央値で 22 分であった。実験を最後まで完了した 401 名の被験者には 400 円を支払った。

401 名の被験者から得られたデータから、我々が用意した検索システム以外を用いて検索を実行した被験者、検索タスクに費やす時間が極端に短すぎる、あるいは長すぎる被験者、タスク後の自由記述で検索結果が操作されていると明示的に記述した被験者を分析データから除いた。さらに、検索タスク中で制御文書を実際に閲覧した検索タスクのデータのみを分析対象とするため、検索タスク中に制御文書をクリックしなかった検索タスクに関するデータを除いた。これらの操作により、最終的に 275 名 (うち、男性 120 名、女性は 149 名、未回答 7 名) の被験者から計 822 件の検索タスクに関するデータを得て分析に用いた。

5 分析

本章では、1 章で述べた 3 つの研究課題それぞれについて分析した結果を報告する。

5.1 RQ1: 文書の意見・信憑性と検索行動との関係

5.1.1 分析対象とした指標

3.1 節で述べた批判的検索行動に関連して、以下の行動を検索タスクごとに測定、分析に用いた。

表 2 意見の違いによる被験者の検索行動の違い。統計的に有意な指標のみ表に示している。

	相反		一致	
	M	SD	M	SD
クエリ投入数	2.67	2.08	2.40	1.77
クリック文書数	6.15	3.51	5.79	3.05
証拠文書数	2.58	1.66	2.33	1.39

表 3 信憑性の違いによる被験者の検索行動の違い。統計的に有意な指標のみ表に示している。

	高信憑性		低信憑性	
	M	SD	M	SD
クリック文書の最低順位	11.01	8.82	12.19	10.72

クエリ投入数 検索タスク中に投入したクエリの数。

クリック文書数 検索タスク中にクリックした文書の数 (制御文書を含む)。

事前の信念と一致した文書のクリック数 検索タスク中にクリックした文書 (制御文書は除く) でユーザの事前の信念と一致した意見を含んでいたものの数。

事前の信念と相反した文書のクリック数 上記と同様に相反した意見を含んでいた文書の数。

クリック文書の最低順位 検索タスク中にクリックした文書のうち、最も下位にランキングされていた文書の順位。

タスク時間 検索タスクの開始から終了までに費やした時間。

ページ閲覧時間 1 つの文書あたりの閲覧時間の平均。

証拠文書数 証拠としてあげた文書の数。

なお、事前の信念と一致 (相反) した文書のクリック数を測定するためには各文書の意見、すなわち、検索タスクについて支持する内容を含むか支持しない内容を含むかに関する情報が必要である。そのため、評価者 1 名 (情報学を専攻する学部生) を雇い、被験者が検索タスク中にクリックした文書集合 (1,103 件) に対して、文書の意見をアノテーションしてもらった。

5.1.2 分析方法

4.4 節で述べたように、本研究では検索タスク中に制御文書をクリックした検索タスクのデータのみを分析に用いた。そのため、分析対象となる 822 件の検索タスクに関するデータは、実験計画は Graco-Latin 方格に基づいていたものの、均衡なデータではない。そこで、分析方法として不均衡なデータに対しても適用できる、線形混合効果回帰分析 (linear mixed-effect regression analysis) を用いた⁵。

具体的には、5.1.1 節であげた各指標 b^i (例: クエリ投入数) について、以下のプロセスに従って b^i を従属変数とする回帰モデルを構築し結果を分析した。まず、被験者および検索タスクを変量効果 (random effects) として扱った帰無モデル m_0^i を構築する。つぎに、 m_0^i に意見を固定効果 (fixed effects) として追加したモデル m_1^i 、 m_1^i に信憑性をさらに固定効果として追加したモデル m_2^i 、 m_2^i に意見と信憑性の交互作用を固定効果として追加したモデル m_3^i という 3 つの回帰モデルを構築する。その後、各モデル $m_j^i (j = 1, 2, 3)$ の適合度が変数を追加する前

3 : <https://azure.microsoft.com/en-us/services/cognitive-services/bing-web-search-api/>

4 : <https://lancers.jp>

5 : 分析ツールとしては R の lme4 パッケージ [1] を用いた。

のモデル m_{j-1}^i よりも有意に改善するかを尤度比検定により評価し、統計的に有意なモデルから得られた結果を報告する。有意水準には 5%を用い、タスク時間およびページ閲覧時間に関しては対数変換をした後、モデルを構築した。

5.1.3 結果

線形混合効果回帰分析の結果、意見の主効果が、クエリ投入数 ($\beta = 0.12, SE = 0.05, t = 2.60, p < 0.01$)、文書クリック数 ($\beta = 0.15, SE = 0.07, t = 2.04, p < 0.05$)、証拠文書数 ($\beta = 0.15, SE = 0.07, t = 2.04, p < 0.05$) に対して認められた。表 2 にこれらの指標を意見の種類ごとにまとめたものを示す。表 2 より、被験者の事前の信念と相反する制御文書を提示したとき、被験者はより多くのクエリを投入し (相反: $M = 2.67$, 一致: $M = 2.40$)、より多くの文書をクリックし (相反: $M = 6.15$, 一致: $M = 5.79$)、より多くの文書を証拠として考慮した (相反: $M = 2.58$, 一致: $M = 2.33$) ことが分かった。

また、信憑性の主効果が、クリック文書の最低順位 ($\beta = -0.49, SE = 0.25, t = -1.99, p < 0.05$) に認められた。表 3 にこの指標を信憑性の違いごとにまとめたものを示す。表 3 より、低信憑性な制御文書を提示したとき、被験者はより下位の文書まで閲覧した (高信憑性: $M = 11.01$, 低信憑性: $M = 12.29$)、ことが分かった。

意見と信憑性の交互作用に関しては、どの指標に対しても、統計的に有意な効果は認められなかった。本研究の分析からは、文書の意見と信憑性の異なる組み合わせが、被験者の検索行動に異なる影響を与えることを示唆する結果は得られなかった。

5.1.4 要約

本節の分析で得られた知見をまとめると以下になる。まず、事前の信念に相反する文書を提示したとき、被験者は検索プロセスより労力を割いたことが分かった。具体的には、そのような被験者はより多くのクエリを投入し、より多くの文書をクリックし、また、より多くの文書を証拠としてあげたことが分かった。逆にいえば、事前の信念と一致する文書を提示したときは、そのような行動をとらないことが分かった。このことは、事前の信念と相反する文書を検索結果中に見せることで、ユーザがより検索プロセスに労力を割り、3.1 節で述べた批判的検索行動をとるようになることを示唆していると考えられる。

5.2 RQ2: 信念の変化

5.2.1 分析方法

文書の意見と信憑性が検索タスク開始前の被験者の信念 (事前の信念) から検索タスク後の信念 (事後の信念) への変化にどのような影響を与えたのかを分析するため、事前の信念と事後の信念の極性の変化に焦点を当てて分析を行った。

極性の変化は 2 種類に分類できる。1 つは、被験者の信念の極性が変化するとき、すなわち、事前の信念として「そう思わない」あるいは「どちらかというと思わない」と回答した被験者が事後の信念として「どちらかというと思う」あるいは「そう思う」と回答 (もしくはその反対) した場合である。もう 1 つは、信念の極性を保持したとき、すなわち、事前の信

表 4 実験条件ごとの信念の変化の分布。また、各事前の信念について、より多くの被験者が事後の信念として回答した側の極性を灰色でハイライトしている。

(a) 相反, 高信憑性		事後の信念			
信念の 極性の変化		No (n=21)	Lean-No (n=49)	Lean-Yes (n=106)	Yes (n=37)
	No (n=32)	15.62%	31.25%	28.12%	25.00%
	Lean No (n=92)	5.43%	21.74%	53.26%	19.57%
	Lean Yes (n=80)	13.75%	22.50%	51.25%	12.50%
	Yes (n=9)	-	11.11%	77.78%	11.11%
(b) 相反, 低信憑性		事後の信念			
信念の 極性の変化		No (n=28)	Lean No (n=35)	Lean Yes (n=101)	Yes (n=30)
	No (n=31)	25.81%	22.58%	41.94%	9.68%
	Lean No (n=101)	10.89%	13.86%	56.44%	18.81%
	Lean Yes (n=57)	14.04%	24.56%	50.88%	10.53%
	Yes (n=5)	20.00%	-	40.00%	40.00%
(c) 一致, 高信憑性		事後の信念			
信念の 極性の変化		No (n=53)	Lean No (n=65)	Lean Yes (n=66)	Yes (n=31)
	No (n=47)	51.06%	29.79%	14.89%	4.26%
	Lean No (n=97)	25.77%	46.39%	26.80%	1.03%
	Lean Yes (n=66)	6.06%	9.09%	48.48%	36.36%
	Yes (n=5)	-	-	20.00%	80.00%
(d) 一致, 低信憑性		事後の信念			
信念の 極性の変化		No (n=37)	Lean No (n=58)	Lean Yes (n=74)	Yes (n=31)
	No (n=37)	37.84%	37.84%	24.32%	-
	Lean No (n=101)	20.79%	46.65%	37.62%	6.93%
	Lean Yes (n=60)	3.33%	15.00%	45.00%	36.67%
	Yes (n=2)	-	-	-	100.00%

念として「そう思わない」あるいは「どちらかというと思わない」と回答した被験者が事後の信念としても「そう思わない」あるいは「どちらかというと思わない」と回答 (もしくはその反対) した場合である。本節では、この 2 種類の信念の変化と制御文書の意見と信憑性との関係について分析する。

分析方法としては、ロジスティック混合効果回帰分析 (logistic mixed-effect regression analysis) を用いて、意見と信憑性の各要因が、信念の極性の変化に影響を与えたかどうかを分析した。具体的には、信念の変化 d_i を 2 値の従属変数 (0: 信念の極性を保持, 1: 信念の極性を変化) とし、意見、信憑性、およびその交互作用を考慮した 3 つの回帰モデルを 5.1 節と同様に構築し、各モデルの適合度が統計的に有意であるかどうかを評価した。

5.2.2 結果

ロジスティック混合回帰分析の結果、意見 ($\beta = 0.65, SE = 0.08, z = 8.53, p < 0.001$) と信憑性 ($\beta = -0.19, SE = 0.08, z = -2.52, p < 0.05$) それぞれに信念の変化に対する主効果が認められた。それぞれの影響を詳しく分析するため、表 4 に実験条件ごとに事前の信念と事後の信念の分布をまとめたものを示す。表中の “No”, “Lean No”, “Lean Yes”, “Yes” は、4.2 節で述べた、信念の測定に用いた質問の回答項目、「そう思わない」、「どちらかというと思わない」、「どちらかというと思う」、「そう思う」、にそれぞれ対応している。たとえば、表 4(a) 中の左上セルにある 15.62% という数値は、相反かつ高信憑性な文書を検索結果に含めた実験条件において、事前

表 5 検索タスクごとに被験者がクリックした文書の意見の分布（制御文書は除く）。表中のタスク ID は表 1 中の ID に対応している。

タスク ID	支持	非支持	両方
1	87	9	5
2	52	5	1
3	52	1	2
4	33	21	3

の信念として「そう思わない」と回答した被験者の 15.62% が事後の信念としても「そう思わない」と回答したこと、つまり、信念を保持したことを表している。

まず、検索タスクの前後で極性を変化させた場合について分析した。表 4 より、事前の信念と相反する文書を提示した場合（①）と一致する文書を提示した場合（②）を比較すると、相反する文書を提示した場合の方が、より多くの被験者が信念を反対の極性へと変化させていたことが分かる。

つぎに、被験者が事前の信念を保持した場合について分析した。表 4 より、事前の信念と一致する文書を提示した場合（③と④）、相反する文書を提示した場合と比べてより多くの被験者が検索タスク後も信念を保持したことが分かった。被験者が信念を保持した場合についてさらに分析すると、高信憑性の文書を提示した場合（③）の方が、低信憑性の文書を提示した場合（④）に比べて、信念を保持した被験者が多いことが分かった。

最後に、表 4(a) および (b) から分かるように、事前の信念と相反する文書を提示した場合、多くの被験者が事後の信念として検索タスクを支持する側に偏っていることが分かる。この現象を理解するため、我々は被験者がクリックした文書の意見の分布を調べた。表 5 は被験者が検索タスク中にクリックした文書（制御文書を除く）が検索タスクについてどのような意見を含んでいたのかの分布である。表 5 より、多くの文書が検索タスクについて支持する内容であったことが分かる。このことは、今回の実験で被験者が検索し閲覧した検索結果の多くは検索タスクを支持する内容であったことを示している。検索結果中の大部分の文書が検索タスクを支持する内容であったため、多くの被験者が検索を通して検索タスクを支持する意見へと信念と変化させたのではないかと考えられる。また、White [16] は自身の研究において検索ユーザの多くは、それ自体が正しい、誤っているに関わらず、検索を通してある治療方法は正しいという事後の信念を抱くというバイアスの存在を報告している。こうしたバイアスも、多くの被験者が事後の信念として支持側に偏ったのではないかと考えられる。一方、注目すべき点として、事前の信念と一致する制御を提示した場合は、通常の実験結果の多くが支持側の内容であったにも関わらずそのような傾向は見られていない（②）。これは、事前の信念と一致する文書を上位に提示することによって確証バイアスが強く働き、被験者の事前の信念を強める方向に作用したのではないかと考えられる。

5.2.3 要 約

本節の分析で得られた知見をまとめると以下になる。被験者の信念の極性の変化は、文書の意見に影響を受けることが分かった。具体的には、事前の信念と相反する制御文書を提

表 6 信念の極性の変化と検索行動の関係。統計的に有意な指標のみ表に示している。

	信念の変化			
	変化		保持	
	<i>M</i>	<i>SD</i>	<i>M</i>	<i>SD</i>
事前の信念と一致した文書のクリック数	0.72	1.37	2.00	2.23
事前の信念と相反した文書のクリック数	3.28	2.83	1.71	2.17

示した場合、より多くの被験者が検索後に信念の極性を変化させ、そうでない場合は事前の信念を保持することが分かった。また、事前の信念を保持する度合いは、文書の信憑性に影響を受け、高信憑性文書を提示した場合の方が、信念を保持する被験者が多いことが分かった。

5.3 RQ3: 意見の変化と検索行動との関係

5.3.1 分析方法

最後に、5.2 節で分析した意見の変化と、5.1 節で分析した検索行動との関係について分析した。両者の関係を分析するため、ロジスティック混合効果回帰分析を用いた。具体的には、5.1.1 節であげた各指標 b^i について、5.2 節と同様に信念の変化 d を 2 値の従属変数とし、 b^i を固定効果、被験者とタスクを変量効果としたロジスティック混合効果回帰モデル M_1^i を構築した。そして、 M_1^i の適合度が被験者とタスクの変量効果のみを考慮した帰無モデル M_0^i に対して統計的に有意に改善するかを尤度比検定により評価した。

5.3.2 結 果

ロジスティック混合効果回帰分析により統計的に有意な主効果が認められた指標の結果を表 6 に示す。事前の信念と一致した文書のクリック数 ($\beta = -0.46, SE = 0.06, z = -7.81, p < 0.001$) および事前の信念と相反する文書のクリック数 ($\beta = 0.26, SE = 0.04, z = 7.44, p < 0.001$) の信念の変化に対する主効果がそれぞれ認められた。表 6 が示すとおり、信念を保持した被験者は検索中に事前の信念と一致する文書を多くクリックする（保持: $M = 2.00$, 変化: $M = 0.72$ ）一方で、信念を変化させた被験者は事前の信念と相反する文書を多くクリックしていた（保持: $M = 1.71$, 変化: $M = 3.28$ ）ことが分かる。

この分析の興味深い点は、事前の信念と相反する制御文書を提示した場合、被験者はクエリの投入数や文書のクリック数などより検索プロセスに労力を割いていたということが分かったにもかかわらず（5.1 節）、そのような行動の違いは信念を変化させたかどうかという観点では得られなかったという点である。これは、被験者が信念を変化させるかどうか、という点と、検索プロセスに労力をかけて情報を調べるかどうか、という点は関連性が少ないのではないかとすることを示唆していると考えられる。

6 議 論

本章では、前章の分析で得られた知見を基に、ユーザが情報をより精査するようなランキングについて議論する。

本研究の大きな知見は、5.1 節で示したように、ユーザの事前の信念と相反する文書を検索結果中に提示することで、ユーザがより積極的に検索プロセスに関与する可能性があるということである。近年ではクエリに対する文書の意見を自動的に分析したり、ユーザの検索ログからそのような意見にユーザがどのように興味を持っているかを推測することは難しくないと考えられる [20]。そのような技術を用いることで、ユーザの事前の信念に相反するような文書をあえて上位にランキングすることで、より慎重な Web 検索を促進することができるのではないかと考えられる。その場合、そのようなランキングにおいてもユーザの満足度を担保するための仕組みについて考える必要がある。

また、信憑性に関していえば、5.2 節の結果から、信憑性の高い文書を閲覧した場合は事前の信念を保持しやすいことが分かった。この結果は、事前の信念と一致した意見を持つ信憑性の高いページに出会うと、自身の確証バイアスをより強く影響を受けてしまい、たとえそのページの内容が不正確でも信じてしまう可能性を示唆している。たとえば、近年の商用検索エンジンはページの表示時間のようなユーザビリティに関する指標をランキングに考慮していると言われている⁶。ユーザビリティの高いページは信憑性が高いと判断されやすく（流暢性バイアスの 1 種として知られる）、洗練されたデザインであったり使いやすいページを上位にランキングすることは、ユーザの批判的な検索行動を抑制し、確証バイアスをより強めてしまう可能性があること示唆している。

7 ま と め

本研究では文書の意見と信憑性がユーザの検索行動と信念の変化に与える影響を分析した。実験の結果、被験者の事前の信念と相反する文書を提示した場合、被験者はより多くのクエリを投入し、より多くの文書をクリックし、そしてより多くの文書を意思決定の証拠としてあげたことが分かった。また、被験者の事前の信念と相反する文書を提示した場合は被験者は信念を反対側の意見へと変化させる可能性が高く、一方で信念と一致する文書を提示した場合は事前の信念をそのまま保持する可能性が高いことが分かった。さらに、そうした信念の変化と検索行動の関連は、クリックする文書の意見の分布以外に差異は見られなかった。

本研究では、被験者が実際に制御文書をクリックした検索タスクのみを分析した。しかし、ユーザは検索結果ページのみからでも影響を受けることがあることが知られている [2]。今後は、被験者が制御文書をクリックしなかった検索タスクのデータも合わせて分析することで、文書の意見や信憑性がどのように影響を与えたのかをより詳しく分析したい。

謝辞 本研究の一部は、JSPS 科学研究費助成事業 18H03494, 18H03243, 18H03244, 16H01756, 16H02906 によるものです。ここに記して謝意を表します。

- [1] D. Bates, M. Mächler, B. Bolker, and S. Walker. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software*, 67(1):1–48, 2015.
- [2] C. L. A. Clarke, E. Agichtein, S. Dumais, and R. W. White. The influence of caption features on clickthrough patterns in web search. In *SIGIR*, pages 135–142, 2007.
- [3] R. H. Ennis. *A taxonomy of critical thinking dispositions and abilities*. W H Freeman/Times Books/Henry Holt & Co, 1987.
- [4] L. Festinger. *Conflict, decision, and dissonance*. Stanford University Press, 1964.
- [5] B. J. Fogg. *Persuasive Technology Using Computers to Change What We Think and Do*. Morgan Kaufmann Publishers, 2003.
- [6] D. Frey and D. Stahlberg. Selection of information after receiving more or less reliable self-threatening information. *Personality and Social Psychology Bulletin*, 12(4):434–441, 1986.
- [7] Y. Kammerer, I. Bråten, P. Gerjets, and H. I. Strømsø. The role of internet-specific epistemic beliefs in laypersons' source evaluations and decisions during web search on a medical issue. *Computers in Human Behavior*, 29(3):1193–1203, 2013.
- [8] D. Kelly. Methods for evaluating interactive information retrieval systems with users. *Foundations and Trends® in Information Retrieval*, 3(1–2):1–224, 2009.
- [9] V. Q. Liao, W.-T. Fu, and S. S. Mamidi. It is all about perspective: an exploration of mitigating selective exposure with aspect indicators. In *CHI*, pages 1439–1448, 2015.
- [10] A. Lowin. Further evidence for an approach-avoidance interpretation of selective exposure. *Journal of Experimental Social Psychology*, 5(3):265 – 271, 1969.
- [11] E. Pariser. *The Filter Bubble: What is the Internet Hiding from You?* Penguin Press, 2011.
- [12] F. A. Pogacar, A. Ghenai, M. D. Smucker, and C. L. Clarke. The positive and negative influence of search results on people's decisions about the efficacy of medical treatments. In *ICTIR*, pages 209–216, 2017.
- [13] P. Resnick, K. R. Garrett, T. Kriplean, S. A. Munson, and N. J. Stroud. Bursting your (filter) bubble: strategies for promoting diverse exposure. In *CWSW*, pages 95–100, 2013.
- [14] C. Schwind, J. Buder, U. Cress, and F. W. Hesse. Preference-inconsistent recommendations: an effective approach for reducing confirmation bias and stimulating divergent thinking? *Computers & Education*, 58:787–796, 2012.
- [15] R. W. White. Beliefs and biases in web search. In *SIGIR*, pages 3–12, 2013.
- [16] R. W. White. Belief dynamics in web search. *Journal of the Association for Information Science and Technology*, 65(11):2165–2178, 2014.
- [17] T. Yamamoto, Y. Yamamoto, and S. Fujita. Exploring people's attitudes and behaviors toward careful information seeking in web search. In *CIKM*, pages 963–972, 2018.
- [18] Y. Yamamoto and T. Yamamoto. Query priming for promoting critical thinking in web search. In *CHIIR*, pages 12–21, 2018.
- [19] Y. Yamamoto, T. Yamamoto, H. Ohshima, and H. Kawakami. Web access literacy scale to evaluate how critically users can browse and search for web information. In *WebSci*, pages 97–106, 2018.
- [20] E. Yom-Tov, S. Dumais, and Q. Guo. Promoting civil discourse through search engine diversity. *Social Science Computer Review*, 32(2):145–154, 2014.

6 : <https://webmasters.googleblog.com/2010/04/using-site-speed-in-web-search-ranking.html>