多様なアルゴリズミック・リコースがもたらす主観的な利益とコストのトレードオフ

冨永 登夢[1,a)]　山下 直美[2,†1]　倉島 健[1]

概要：アルゴリズミック・リコースとは，ユーザが AI システムから下された不利な判定を覆すための反事実に基づく行動計画を指す．多様な (複数の) リコースが提供された場合，ユーザの判定に対する理解度や行動計画に対する実行意欲は向上し得るが，反事実的思考に由来する認知的負担や感情的負荷も同時に増大しかねない．このトレードオフを明らかにするため，本研究は自動車ローン審査を再現する被験者間計画の統制実験 (N=750) を実施した．我々は，特定の数 (1, 3, 7) の反実仮想サンプルを以下のポリシーで選ぶことで，リコース集合の多様性を実験的に操作した：近接性のみ考慮する Close 集合，近接性と相互差異性の双方を考慮する Diverse 集合．結果として，リコース数が 3 の時，Diverse 集合は Close 集合と比較して主観的利益，特に実行意欲，を向上させた．また，Diverse 集合において，主観的利益はリコース数が 3 と 7 で同等であったが，認知的負担はリコース数に比例して増加した．定性的分析から，Diverse 集合の被験者は，有意義で心理的に抵抗感のないプランを見つけられたことが分かった．これらの結果は，主観的な利益とコストのバランスは少数の多様なリコース集合により最適化されることを示唆する．

1. はじめに

与信や医療などの高リスク領域において AI モデルの運用が進むなか，“説明を受ける権利” を保証する有力なアプローチとして，アルゴリズミック・リコースが注目されている．リコースは，ユーザが自身にとって不利な AI 判定を覆せるように実行可能な行動案を反実仮想説明の形式で提示する [13], [34]．例えば，AI システムによりローン申請が不採用となった申請者に対して，“もしあなたの年収が$1,000 高ければ承認されます” という説明がこれに該当する．リコースは，説明として理解しやすく，行動計画として実行に移しやすいものであることが望ましい [12]．

これを実現する有望なアプローチの 1 つにリコースの多様化がある．これは，特徴量空間において相互の非類似性を最大化する複数の反実仮想サンプルを選ぶ方法を指す [25]．ここで，反実仮想サンプルとは対象ユーザと異なる判定を受けた別のユーザを指す．最終的に，各サンプルから構成されたリコースが 1 つの集合として提示される．このように多様化されたリコース集合は，複数の多角的な説明と行動選択肢を提供することで，AI システムの解釈性の向上 [30], [38] や，現実の制約下でも実行できるプランの提案 [1] など，様々な利益をユーザに与える可能性を持つ．一方で，反事実的思考にともなう認知負荷 [4], [5] や，後悔や罪悪感といった負の感情経験 [6] などの心理的なコストが多様化によって増幅することも同時に懸念される．このような利益とコストのトレードオフは理論的には予想されるが，実証的にはこれまで検討されてこなかった．そのため，実応用において両者のバランスを最適化するためにリコース集合の多様性をどう制御すべきかは未知である．

この問題に対処するため，本研究は自動車ローン申請を再現した被験者間計画のユーザ実験 (N=750) を実施した．実験では，大きさ 1，3，または 7 のリコース集合を (1) ユーザに最も近接する反実仮想サンプルを選ぶ方法 (Close)，もしくは (2) ユーザに近接するが互いに大きく相違する反実仮想サンプルを選ぶ方法 (Diverse) で構成することで，リコースの数と多様性を操作した．Close と Diverse の概念的な違いを図 1 に示す．これらのリコース集合に対して，被験者は利益の指標として説明合理性，実行可能性，行動意欲，および判定受容を，コストの指標として認知的負荷と感情的負担を評価した．本実験を通じて我々は，**リコースの多様性と数は，ユーザが認識する利益とコストにどのような影響を及ぼすか？**という研究課題に取り組んだ．

本実験から，主観的な利益とコストはそれぞれ対照的なパターンを示すことが確認された．主観的利益について，Diverse 集合は 1 件から 3 件にかけて増加を示したが，そ

[1] NTT(株) 人間情報研究所
[2] NTT(株) コミュニケーション科学基礎研究所
[†1] 現在，京都大学大学院情報学研究科
[a)] tomu.tominaga@ntt.com

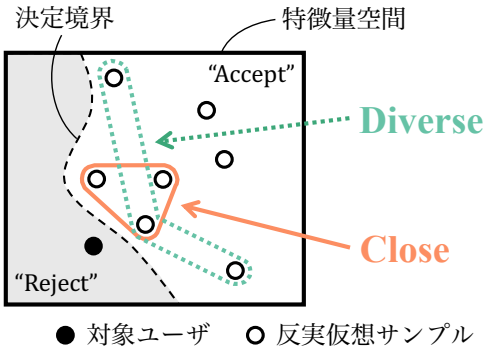

図 1 Close と Diverse の反実仮想サンプル選択ポリシー

れ以降は頭打ちとなった．一方，Close 集合はリコース数に対して段階的な増加を示し，7 件の時に Diverse 集合と同等となった．この傾向は，特に行動意欲において顕著であった．主観的コストについては，Close 集合では 1 件から 3 件にかけて認知負荷が上昇したのち安定したのに対し，Diverse 集合ではリコース数に応じて漸増し，3 件時点では Close 集合と同等だったが，7 件では Close 集合を上回った．定性分析から，3 件の多様な選択肢があると，被験者は自分にとって有意義で，必要で，心理的に受け入れられる計画を見いだしやすく，そのことが行動意欲を強めた一方，7 件になると意思決定基準や提案行動の不確実性への懸念が高まり，結果として Close 集合より大きな認知的労力を要した可能性が示唆された．

本研究には 3 つの貢献がある．まず，リコースの多様化による利益とコストのトレードオフ関係ついて実証的知見を提供した．次に，両者が最適化される条件として，比較的小規模のリコース集合の多様化が心理的な負荷を抑えて実行意欲を高めることを確認した．さらに，これらの知見にもとづき，リコース集合の効用を最大化するために多様性を制御する方法を議論した．

2. 関連研究

アルゴリズミック・リコースの目的は，不利な結果を受けたユーザに対し，理由の明確化と実行可能な手立ての提示を通じて支援することにある [34], [35]．形式的には，ユーザは言及される原因が少ない説明を好み [24]，必要労力の小さい行動計画を好む [13], [34] という仮定のもと，N 次元の特徴空間を $\mathcal{X} = \mathcal{X}_1 \times \cdots \times \mathcal{X}_N$，二値分類器を $f : \mathcal{X} \to \{+1, -1\}$，距離関数を $\mathrm{dist} : \mathcal{X} \times \mathcal{X} \to \mathbb{R}_{\geq 0}$ として，正例領域 $\mathcal{A} = \{\, \boldsymbol{c} \in \mathcal{X} \mid f(\boldsymbol{c}) \neq f(\boldsymbol{x}) \,\}$ から $f(\boldsymbol{x}) = -1$ である対象ユーザ $\boldsymbol{x} \in \mathcal{X}$ に最も近い反実仮想サンプル $\boldsymbol{c}^*$ を特定する以下の最適化問題を解き，

$$\boldsymbol{c}^* \in \arg\min_{\boldsymbol{c} \in \mathcal{A}} \mathrm{dist}(\boldsymbol{c}, \boldsymbol{x}), \tag{1}$$

リコースを摂動 $\boldsymbol{\delta} = \boldsymbol{c}^* - \boldsymbol{x}$ として導く．これを基本する技術的拡張が現在 300 以上考案されている [12], [36]．

この中で我々はリコースの多様化に注目する．これは，互いにできるだけ大きく異なる複数の反実仮想サンプルを選ぶことで多様なリコース集合を生成する方法である [25]．いくつかの技術 [18] の中でも，代表例として，Mothilal らは，目的関数に Determinantal Point Process(DPP) 項を加えてリコース間の相互多様性を促す方法 DiCE [25] を構築した．彼らは多様性を定量的に評価する指標も考案し，それに基づいてベンチマークで検証した結果，DiCE が先行手法に匹敵する近接性を保ちつつ，より多様なリコースを導出できることを示した [25]．

本研究は，このように生成された多様なリコース集合に対しユーザがどのような反応を示すかに関心がある．リコースの多様化が説明や推薦としてユーザにどのような利益をもたらすかは既存研究から予測される．たとえば説明の観点では，複数の説明は複雑な概念に対する理解を促すことから [30], [38]，AI システムの判定基準への理解度向上が期待される．実際，反実仮想説明が 1 つだけ与えられた群より，複数与えられた群の方が AI に関する理解度テストで高成績だったと報告されている [3]．理解度の向上は不利な結果の受容につながるため [32]，多様なリコースはより説得力のある説明として機能しうる．また，行動計画の推薦としては，さらに二つの利点が見込まれる．第一に，選択肢の幅が広がることで，ユーザは現実の制約下において適合する行動を選びやすくなる [1], [18], [37]．第二に，自己決定理論 [28] に論じられているように，複数の選択肢を比較し最適と判断したものを選ぶ行為自体が主体性を強め，実行意欲を喚起する可能性がある．

しかし，多様なリコースは認知的および感情的コストをともなうことも同時に予想される．反事実的思考は，“もし原因が X ではなく X' であれば，結果は Y ではなく Y' である” のように，二重の因果性を同時に検討する必要がある [21] ため，因果的思考より認知負荷が高い [4], [5]．また，現実と理想を照らし合わせることで，過去に別の行動をしていれば結果がどうなったかを想像して後悔を抱いたり [26], [40]，罪悪感や自己非難といった否定的感情を持ったりすることにつながる [6]．近年の研究では，リコースに過去の不足を非難されているかのように感じたと報告する参加者も確認されている [32]．そのため，多様化により多種多様な反実仮想に曝されることによる，こうした心理的コストの増大が懸念される．

本研究は，このようなリコースの多様化による利益とコストのトレードオフに関する実証的知見を獲得し，実用的にリコース多様性を制御する最適条件を明らかにする．

3. 実験

我々は，リコースの多様性がユーザの反応に及ぼす影響を検証するため，自動車ローン審査を再現するシナリオのもと 750 名の被験者を対象に被験者間計画の統制実験を

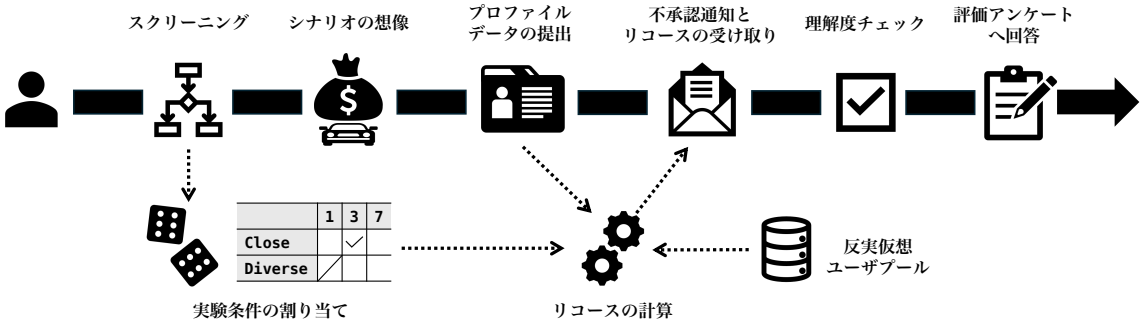

図 2 実験の流れの全体像

2025 年 2 月から 3 月にオンラインで実施した．手続き全体を図 2 に示す．なお，本実験は公益財団法人パブリックヘルスリサーチセンターの倫理審査委員会から承認を受けたものである (承認番号：PHRF-IRB 25A0004)．

3.1 被験者

我々はオンライン調査会社を通じて被験者を募集した．実験参加条件は，(1) 民間企業または公的機関に勤務，(2) 自動車購入を検討中，(3) 参加時点でローン未保有，(4) 年収 1,000 万円未満，とした．実験参加条件を満たす被験者は，実験内容に関する説明を受けた後に，同意確認を提出した．この時点で，1062 名が被験者として実験に参加した (女性 300 名，男性 762 名，平均年齢 49.2 歳)．同意確認後，各被験者に実験条件を無作為に割り当てた．実験終了後，被験者は 600 円相当の謝礼を受け取った．

3.2 シナリオ

本実験では，先行研究 [31], [32] を参考に自動車ローン申請のシナリオを採用した．具体的な内容を以下に示す．

> あなたは，自身の年収の 3 分の 1 に相当する 2 年の自動車ローンを借りたい．そこで金融機関を訪れ，審査用のプロファイルデータを提出した．AI システムによる審査の結果，今回のローン申請は不承認という結果となった．あなたは，不承認となったの理由と，承認されるために必要な行動を知るため，AI が生成した複数のプランを確認することにした．

ここでは，上記金融機関 (AI) は年収の 1/4 を超える融資申請を全て却下するという内部規則にしたがって審査すると仮定した．そのため，全ての被験者の申請は却下されるが，この方針は実験中に被験者には開示されない．

3.3 プロファイルデータの提出

既存研究 [31], [32] を参考に選定した審査用のプロファイル項目を表 1 に示す．これらは，基本属性 (#1–#3)，現職情報 (#4–#10)，職歴と技能 (#11–#14)，人的ネットワーク (#15,#16) に分類される．なお，性別，出生地，国籍など個人の努力で変更できない属性は含めていない [14]．

3.4 実験条件

プロファイルデータが提出された後，我々は被験者ごとに割り当てられた実験条件に基づいてリコース集合を導出した．具体的な計算方法は 3.5 節で述べる．ここでは，本実験で操作した実験条件の詳細を説明する．

3.4.1 多様性基準：Close もしくは Diverse

我々は，対象ユーザに最も近接する反実仮想サンプルを選ぶ方法 [37]，もしくは対象ユーザへの近接性を保ちつつ相互の非類似性が最大となるよう選ぶ方法 [25] のいずれかでリコース集合を構成することでその多様性を操作した．これらをそれぞれ Close 集合，Diverse 集合と呼ぶ．

Close 集合では近接性のみで選定するため，変更量が小さいリコースが集合に含まれやすい．一方，Diverse 集合は近接性とサンプル間非類似性の結合スコアを最大化するため，対象ユーザへの近接性を保ちながら，幅広いばらつきを持つリコースを含めることができる．

3.4.2 選択肢数：1 件，3 件，もしくは 7 件

リコース集合に含まれる選択肢数は 1 件，3 件，7 件の 3 水準とした．単一選択肢は，複数選択肢の効果を評価するための基準条件として用いた．なお，選択肢 1 件では多様性は存在し得ないため，1 件の Diverse 集合は考慮しない．

複数選択肢の最小件数として 3 件を採用した．選択肢が二者のみの時，類似判断は “似ているか否か” という二値比較に留まりやすいが，三者以上では “A と B は似ているが C は異なる”，“すべて異なる” のように認知プロセスが分類やグルーピングに発展し，多次元的な類似判断と多様性の知覚が促進される [22]．以上を踏まえ，本研究では多様性の認識に必要な最小件数として 3 を用いた．

また，選択肢の上限値として 7 件を設定した．短期記憶容量に関して古典的には “7±2” に保持や比較の限界があり [23]，また，近年の研究では選択肢過多は一部の無視，無作為な選択，選択の回避を招きうることも報告されている [29]．よって、7 件以上増やしても差異が観察されづらいと想定して，本実験では最大件数を 7 とした．

表 1 審査用プロファイルデータ項目 ($\boldsymbol{x}_i$ と $\boldsymbol{x}_i^*$ は入力サンプルと反実仮想サンプルの要素 i)

#	項目 (特徴量)	選択肢 (オプション)	制約条件	変数尺度
1	居住地	1. 東京 / 2. 東京以外		カテゴリ
2	居住形態	1. 持ち家 / 2. 賃貸・寮		カテゴリ
3	最終学歴	1. 高校卒 / 2. 短大/専門卒 / 3. 大学卒 / 4. 院卒 (修士) / 5. 院卒 (博士)	$\boldsymbol{x}_3^* \geq \boldsymbol{x}_3$	順序
4	勤務先	1. 一般企業 / 2. 公的機関		カテゴリ
5	職位	1. 一般社員 / 2. 主任 / 3. 係長 / 4. 課長 / 5. 次長 / 6. 部長 / 7. 本部長・事業部長 / 8. 常務取締役 / 9. 専務取締役 / 10. 代表取締役	$\boldsymbol{x}_5^* \geq \boldsymbol{x}_5$	順序
6	勤続年数 (年)	1. 0-1 / 2. 1-3 / 3. 3-5 / 4. 5-10 / 5. 10-20 / 6. 20-	$\boldsymbol{x}_6^* \geq \boldsymbol{x}_6$	順序
7	マネジメント歴 (年)	1. なし / 2. 0-1 / 3. 1-3 / 4. 3-5 / 5. 5-10 / 6. 10-20 / 7. 20-	$\boldsymbol{x}_7^* \geq \boldsymbol{x}_7$	順序
8	勤務時間 (時間/日)	1. 0-2 / 2. 2-4 / 3. 4-6 / 4. 6-8 / 5. 8-10 / 6. 10-12 / 7. 12-		順序
9	在宅勤務時間 (時間/日)	1. 0-2 / 2. 2-4 / 3. 4-6 / 4. 6-8 / 5. 8-10 / 6. 10-12 / 7. 12-		順序
10	副業数	1. なし / 2. 1 / 3. 2 / 4. 3 / 5. 4 / 6. 5-		順序
11	転職歴	1. なし / 2. あり	$\boldsymbol{x}_{11}^* \geq \boldsymbol{x}_{11}$	順序
12	海外勤務歴	1. なし / 2. あり	$\boldsymbol{x}_{12}^* \geq \boldsymbol{x}_{12}$	順序
13	海外留学歴	1. なし / 2. あり	$\boldsymbol{x}_{13}^* \geq \boldsymbol{x}_{13}$	順序
14	TOEIC 最高得点	1. なし / 2. 10-400 / 3. 400-495 / 4. 500-595 / 5. 600-695 / 6. 700-795 / 7. 800-895 / 8. 900-990	$\boldsymbol{x}_{14}^* \geq \boldsymbol{x}_{14}$	順序
15	Facebook 利用	1. 未登録・友達なし / 2. 登録済・友達あり		カテゴリ
16	LinkedIn 利用	1. 未登録・友達なし / 2. 登録済・友達あり		カテゴリ

3.5 リコース集合の計算

実験条件として被験者に割り当てられた多様性基準を p, 選択肢数を k とした時，被験者に提示されるリコース集合 C^* は候補となる反実仮想サンプルの集合を $C=\{\boldsymbol{c}^1, ..., \boldsymbol{c}^k\}$ として以下の最適化問題を解くことで得られる．

$$C^* = \begin{cases} \underset{\substack{C\subset(\mathcal{A}\cap\mathcal{Z})\\|C|=k}}{\arg\min} \ \dfrac{1}{k}\displaystyle\sum_{\boldsymbol{c}\in C} \mathrm{dist}(\boldsymbol{c},\boldsymbol{x}) & \text{if } p\text{=Close}, \\ \underset{\substack{C\subset(\mathcal{A}\cap\mathcal{Z})\\|C|=k}}{\arg\min} \ \dfrac{1}{k}\displaystyle\sum_{\boldsymbol{c}\in C} \mathrm{dist}(\boldsymbol{c},\boldsymbol{x}) - \dfrac{1}{\binom{k}{2}}\displaystyle\sum_{\substack{\boldsymbol{c}^n,\boldsymbol{c}^m\in C\\ n<m}} \mathrm{dist}\left(\boldsymbol{c}^n,\boldsymbol{c}^m\right) & \text{if } p\text{=Diverse}, \end{cases} \tag{2}$$

候補集合 C は，正例領域 $\mathcal{A}$ (式 (1)) と変更可能領域 $\mathcal{Z}$ の両方に属するよう制約される．本シナリオでは被験者は年収の 3 分の 1 を申請し，金融機関 (AI) は年収の 4 分の 1 を超える申請を拒否するように設計されているため，$\mathcal{A}$ は被験者の年収の 4/3 以上の所得を持つ反実仮想サンプル群に対応する．変更可能領域 $\mathcal{Z}$ は，表 1 に示した特徴量の大小関係の制約を満たすサンプル群であり，学歴やベストスコアを下げるといった非現実的変更を防ぐ．不等号制約をもつ特徴量の ID 集合を $\mathcal{I}_{\geq}$ として，$\mathcal{Z} = \{\boldsymbol{c} \in \mathcal{X} \mid \boldsymbol{c}_i \geq \boldsymbol{x}_i,\ i \in \mathcal{I}_{\geq}\}$ と定義される．つまり，C は被験者の 4/3 以上の年収で表 1 の不等式制約を満たす反実仮想サンプル群となる．

距離関数 dist は先行研究 [25], [37] にしたがって特徴ワイズ $L1$ 距離を用いた．カテゴリ尺度の特徴集合を $\mathcal{I}_{\mathrm{cat}}$, 順序尺度の特徴集合を $\mathcal{I}_{\mathrm{ord}}$ とすると，

$$\mathrm{dist}(\boldsymbol{c},\boldsymbol{x}) = \frac{1}{|\mathcal{I}_{\mathrm{cat}}|}\sum_{i\in\mathcal{I}_{\mathrm{cat}}} \mathbf{1}[\,\boldsymbol{c}_i \neq \boldsymbol{x}_i\,] + \frac{1}{|\mathcal{I}_{\mathrm{ord}}|}\sum_{i\in\mathcal{I}_{\mathrm{ord}}} \frac{|\boldsymbol{c}_i - \boldsymbol{x}_i|}{\mathrm{MAD}_i}, \tag{3}$$

表 2 リコース集合の提示例 (リコース数 k=3 の場合)

	被験者	行動プラン		
		A	B	C
居住地	東京	東京以外	東京	東京以外
...	...	...	...	...
職位	主任	主任	係長	課長
...	...	...	...	...
LinkedIn	登録済み	登録済み	登録済み	登録済み

と定義される．ここで $\mathbf{1}[\cdot]$ は条件式が真であれば 1，偽であれば 0 を返す指示関数，MAD_i は第 i 特徴量の中央値絶対偏差を指す．第一項はカテゴリ変数の不一致率，第二項は順序変数の MAD 正規化差分の平均に相当する．

式 (2) の第一項は被験者 $\boldsymbol{x}$ と反実仮想サンプル $\boldsymbol{c}$ の平均距離である．Close 集合はこの距離を最小化し，乖離が小さいリコースを得る．第二項は，先行研究 [25] が考案した反実仮想サンプル間の多様性尺度で，反実仮想サンプル間の相互距離の平均である．これを第一項から減ずることで，Diverse 集合は近接性を保ちながら多様性を最大化する．

以上の計算から導出された k 件のリコースを行動プランの集合として被験者に提示した．表 2 に提示例を示す．行動プランの表示順は無作為に定めた．

3.6 理解度確認タスク

本実験では反実仮想説明の理解に要する認知的負荷を測定するため，被験者がリコースの内容を正確に理解できていることを前提とするのが妥当である．そこで我々は，今回不承認となった原因と次回以降承認されるために必要な行動としてリコースが指摘するプロファイル項目を多岐選

択式で回答させた．回答結果とリコース内容に整合性のない被験者は実験から除外した．最終的な被験者は 750 名となった (男性 534 名，女性 216 名，平均年齢 48.4 歳)．

3.7 リコース集合に対する反応の測定

我々は，リコース集合がユーザに与える利益とコストを主観評価のアンケートで計測した．なお，リコースの理解に要する認知的負荷を精緻に評価するために，理解度の確認の直後に認知的負荷の計測を実施した．その後，被験者は利益に関するアンケート，否定的感情に関するアンケートの順で回答した．

3.7.1 利益の観点

本研究では，リコースによる主観的利益を包括的に評価するために 4 つの観点を定めた．まず，リコース自体の評価観点として (1) **説明合理性**：不承認理由の説明としての妥当性，(2) **実行可能性**：行動計画としての実行しやすさ，(3) **実行意欲**：行動計画に対するやる気，を設けた．単一選択肢条件の被験者は，提示されたリコースを各観点で 7 段階で評価し，評価理由を自由記述式で回答した．複数選択肢条件の被験者は，事前に各観点において最もよいリコースを集合の中から選び，それぞれリコースに対して 7 段階評価と評価理由の自由記述回答を行った．さらに，両条件の被験者に共通して，AI 判定に対する評価として，(4) **判定受容**：判定結果に対する納得度，を設けた．

3.7.2 コストの観点

リコースの内容を理解するための認知的負荷を観察するため，理解度確認タスク (3.6 節) における認知的負荷を NASA-TLX [11] に基づいて測定した．NASA-TLX は 6 次元から成るが，本研究では認知的負荷に関連すると想定される (1) **精神的要求**：タスク遂行に要した認知的活動，(2) **労力**：タスク遂行に費やした身体的・精神的努力，(3) **フラストレーション**：タスク遂行における苛立ちやストレス，に焦点を当てた．各次元は 0〜100 の 5 刻みで評定した．

我々は，既存研究 [6], [31], [32] を参考に，反実仮想思考が別の過去や個人的な不足を強調することで喚起される (4) **否定的感情**：後悔，恥，罪悪感，自己避難，失望，嫌悪，不満，差別感，を被験者が経験したかどうかを調べた．具体的には，これらの 8 つの感情のうち経験した感情を多岐選択式で回答させ，その選択数を評価した．

3.8 実験条件の操作性確認

我々は，複数リコース条件の参加者に対し，“提示された行動計画全体は，多様な観点から判定結果を説明していると思いますか？” と 7 件法で尋ねた．回答結果に対して二要因分散分析を実施した結果，リコース数と多様性基準の双方で主効果が有意であり (リコース数：$F(1,607)=9.77$, $p=0.002$, $\eta_p^2=0.016$；多様性基準：$F(1,607)=4.15$, $p=0.042$, $\eta_p^2=0.007$)，3 件よりも 7 件で，Close 集合よりも Diverse 集合で多様性が知覚されていることを確認した (Close-3: 3.27, Diverse-3: 3.71, Close-7: 3.80, Diverse-7: 3.87)．

表 3 各実験条件の被験者数

	1 件	3 件	7 件
Close 集合	139	152	168
Diverse 集合	–	126	165

4. 分析

理解度確認を経て最終的に分析対象となった被験者 750 名の実験条件に対する分布を表 3 に示す．我々はリコース集合の多様性の影響を定量分析と定性分析の双方で調べた．

4.1 定量分析

主解析として，主観的な利益とコストの各測定値に対し，多様性基準 (Close/Diverse) とリコース数 (3 件/7 件) を要因とする二要因分散分析を実施した．交互作用効果が有意のときは，Tukey HSD で各条件の事後比較を行った．単一選択肢条件は多様性が定義できないため分散分析から除外し，Close-1 を参照条件とする Dunnett 検定により複数選択肢を提示する効果を評価した．

4.2 定性分析

定量的結果をより深く理解するため，リコースに対する評価 (説明合理性，実行可能性，実行意欲) の理由に関する自由記述をコーディングした．具体的には，まず筆頭著者がデータ全体を反復的にレビューしてコードブックを作成し，評価観点ごとに層化無作為抽出した 10%の回答 (各 75 件，全 225 件) を外部協力者がコードブックにしたがって独立にコーディングした．両者の一致度は Cohen の κ 係数で 0.61〜0.64 となり，実質的な合意が得られた [8], [17]．残りは筆頭著者が同手順でコーディングし，最終的に横断的テーマに統合した．

5. 結果

5.1 定量的結果

5.1.1 主観的利益

図 3 は，説明合理性，実行可能性，実行意欲，および判定受容について，各条件における被験者評価値の分布を示す．説明合理性では，リコース数の主効果のみ有意であった ($F(1,607)=7.85$, $p=0.005$, $\eta_p^2=0.013$)．Dunnett 検定では，他の全条件で説明合理性が高かった (Diverse-3：$p<0.001$, Close-3：$p=0.019$, Diverse-7：$p<0.001$, Close-7：$p<0.001$)．したがって，多様性に関係なく，選択肢が増えると少なくとも 1 つを合理的だと捉える可能性が高まるといえる．

実行可能性では，リコース数の主効果 ($F(1,607)=4.04$,

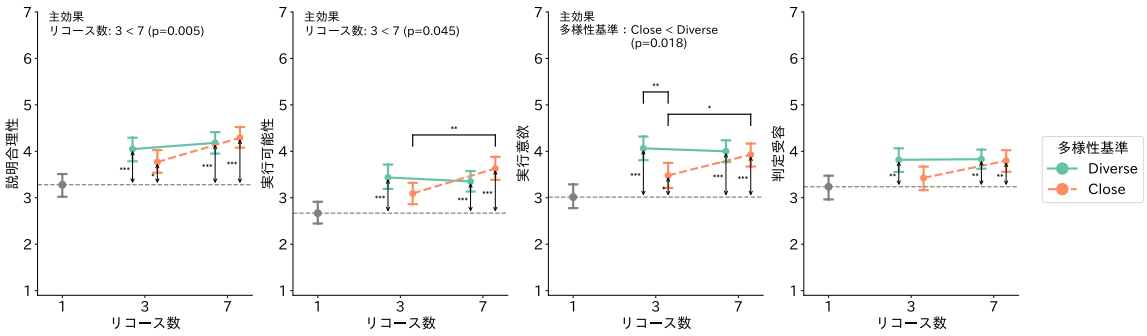

図 3 説明合理性，実行可能性，実行意欲，判定受容に関する二要因分散分析，Tukey HSD，および Close-1(灰色) を基準とした Dunnett の多重比較の結果．いずれも多重比較補正済み (*：p<0.05，**：p<0.01，***：p<0.001)．

p=0.045, η_p^2=0.007) と交互作用 ($F(1,607)$=6.57, p=0.011, η_p^2=0.011) が有意であった．Tukey HSD では，Close において 3 件から 7 件への増加で実行可能性が有意に向上した (p=0.002)．Dunnett 検定では，Diverse-3，Diverse-7，Close-7 が Close-1 より高く (各 p<0.001)，Close-3 は有意差がなかった (p=0.052)．近接性に基づくリコースでは選択肢を増やすことで実行可能な計画を見つけやすくなり，多様性に配慮すると 1 件から複数件への移行で効果が現れ，複数件中での増加は小さいことを示唆する．

実行意欲では，多様性基準の主効果 ($F(1,607)$=5.67, p=0.018, η_p^2=0.009) と交互作用 ($F(1,607)$=4.00, p=0.046, η_p^2=0.007) が有意であった．Tukey HSD では，3 件のときに Close より Diverse が高く (p=0.002)，Close では 3 件よりも 7 件が高かった (p=0.016)．Dunnett 検定では，全条件が Close-1 より高かった (Diverse-3：p<0.001，Close-3：p=0.039，Diverse-7：p<0.001，Close-7：p<0.001)．すなわち，Diverse は選択肢が単一から複数になった段階で実行意欲を押し上げ，その後は横ばいとなる一方，Close では集合規模の大きさにともない段階的に高まるといえる．

判定受容では，有意な主効果および交互作用は認められなかった．ただし Dunnett 検定では，Diverse-3(p=0.007)，Diverse-7(p=0.003)，Close-7(p=0.005) が Close-1 より高く，Close-3 は有意差がなかった (p=0.651)．したがって，複数のリコースが提示されれば受容は概ね高まるが，多様性や数そのものへの依存は小さいといえる．

5.1.2 主観的コスト

図 4 は, 精神的要求, 労力, フラストレーション, および否定的感情経験について, 各条件の分布を示す．精神的要求では, リコース数の主効果 ($F(1,607)$=9.019, p=0.003, η_p^2=0.015) と交互作用 ($F(1,607)$=10.121，p=0.002，η_p^2=0.016) が有意であった．Tukey HSD より，Diverse において 3 件より 7 件が高く (p<0.001)，リコース数 7 件において Close より Diverse が高かった (p=0.009)．Dunnett 検定では，Close-3(p=0.001)，Diverse-7(p<0.001)，Close-7(p=0.001) が高かった．精神的要求は，Close では単一から複数への上昇後に頭打ちとなるが，Diverse では集合の大きさに比例して増加し最終的に Close を上回ることを示唆する．

労力は交互作用のみ有意であった ($F(1,607)$=8.773, p=0.003, η_p^2=0.014)．Tukey HSD より，リコース数 3 件において Close より Diverse が低く (p<0.05)，Diverse において 3 件より 7 件が高い (p<0.001) ことが確認された．Dunnett 検定では，Close-3(p=0.011) と Diverse-7(p<0.001) が高かった．多様化された 3 件は 1 件と同程度の労力で理解できる一方，7 件では Close の複数提示に匹敵する労力が必要となるといえる．

フラストレーションでは，主効果と交互作用効果のいずれも有意でなかった．Dunnett 検定では Diverse-7 のみが基準より高かった (p=0.020)．すなわち，単一から複数への移行でフラストレーションは限定的にしか変化せず，多様性や数による系統的な差は見られないことを示唆する．

否定的感情経験では，多様性基準の主効果のみ有意 ($F(1,607)$=7.949，p=0.005，η_p^2=0.013) であった．Dunnett 検定では，いずれの条件も Close-1 との差は有意でなかった (Close-3：p=0.919，Diverse-3：p=0.158，Close-7：p=1.000，Diverse-7：p=0.094)．つまり，否定的感情は選択肢数には影響されないが，複数提示において多様化が負の感情的経験を緩和する傾向があるといえる．

5.2 定性的結果

自由記述の定性分析から，説明合理性 (R) で 8 コード，実行可能性 (A) で 11 コード，実行意欲 (W) で 9 コードが抽出され，これらは最終的に 4 つのテーマに集約された．結果を表 4 に示す．各コードは R/A/W-#(例：R-1) の形式で参照する．

5.2.1 裁量・外的制約

選択肢が増えるにつれ, 実行可能性・実行意欲の両方で外的制約 (A-1，W-1) の言及が減少し，時間と金銭 (A-2) やスキルと経験 (A-4) も低下した一方で，社会的地位 (A-5) は緩やかに増加した (表 4)．管理可能性 (A-3) と努力の程度 (W-2) も件数の増加にともなって重視される傾向にあり，3

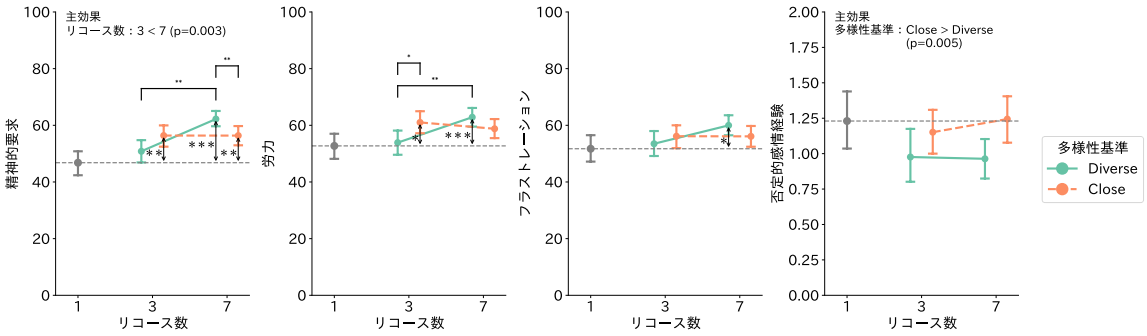

図 4 精神的要求，労力，フラストレーション，否定的感情経験に関する二要因分散分析，Tukey HSD，および Close-1(灰色) を基準とした Dunnett の多重比較の結果．いずれも多重比較補正済み (*：p<0.05，**：p<0.01，***：p<0.001)．

表 4 各評価観点におけるテーマとコードの条件別分布

			1 件	3 件		7 件	
評価観点	テーマ	コード	Close	Close	Diverse	Close	Diverse
説明合理性 (R)	実現性・実用性	1. 実現尤度	13 (11.4%)	21 (16.8%)	17 (18.3%)	17 (13.1%)	18 (14.0%)
	価値観・生活	2. 内省	12 (10.5%)	14 (11.2%)	16 (17.2%)	30 (23.1%)	31 (24.0%)
		3. 消極的受容	9 (7.9%)	7 (5.6%)	7 (7.5%)	9 (6.9%)	5 (3.9%)
	説明責任	4. 妥当性の指摘	49 (43.0%)	51 (40.8%)	36 (38.7%)	51 (39.2%)	50 (38.8%)
		5. 明瞭性欠如	15 (13.2%)	11 (8.8%)	5 (5.4%)	9 (6.9%)	10 (7.8%)
		6. 不確実性とリスク	2 (1.8%)	7 (5.6%)	3 (3.2%)	4 (3.1%)	2 (1.6%)
		7. 不公平性	4 (3.5%)	6 (4.8%)	3 (3.2%)	4 (3.1%)	6 (4.7%)
		8. AI 不信	3 (2.6%)	1 (0.8%)	0 (0.0%)	0 (0.0%)	0 (0.0%)
実行可能性 (A)	裁量・外的制約	1. 外的制約	27 (20.8%)	16 (11.9%)	13 (12.7%)	14 (9.9%)	13 (9.7%)
		2. 時間と金銭	26 (20.0%)	13 (9.7%)	11 (10.8%)	16 (11.3%)	16 (11.9%)
		3. 管理可能性	22 (16.9%)	26 (19.4%)	25 (24.5%)	43 (30.3%)	29 (21.6%)
		4. スキルと経験	14 (10.8%)	14 (10.4%)	6 (5.9%)	14 (9.9%)	9 (6.7%)
		5. 社会的地位	8 (6.2%)	15 (11.2%)	6 (5.9%)	14 (9.9%)	16 (11.9%)
	実現性・実用性	6. 一貫性欠如	1 (0.8%)	0 (0.0%)	1 (1.0%)	0 (0.0%)	0 (0.0%)
		7. 難易度一般	14 (10.8%)	7 (5.2%)	10 (9.8%)	12 (8.5%)	15 (11.2%)
	価値観・生活	8. 心理的抵抗	8 (6.2%)	13 (9.7%)	6 (5.9%)	5 (3.5%)	12 (9.0%)
		9. 実生活の影響	4 (3.1%)	22 (16.4%)	13 (12.7%)	18 (12.7%)	8 (6.0%)
	説明責任	10.不確実性とリスク	4 (3.1%)	3 (2.2%)	8 (7.8%)	2 (1.4%)	11 (8.2%)
		11.AI 忌避	0 (0.0%)	1 (0.7%)	0 (0.0%)	1 (0.7%)	0 (0.0%)
実行意欲 (W)	裁量・外的制約	1. 外的制約	9 (7.5%)	10 (7.6%)	4 (3.7%)	5 (3.4%)	6 (4.3%)
		2. 努力の程度	27 (22.5%)	29 (22.0%)	37 (34.3%)	43 (29.7%)	31 (22.5%)
	実現性・実用性	3. 実現性の判断	24 (20.0%)	36 (27.3%)	30 (27.8%)	40 (27.6%)	39 (28.3%)
		4. 利得	6 (5.0%)	14 (10.6%)	11 (10.2%)	7 (4.8%)	11 (8.0%)
	価値観・生活	5. 心理的抵抗感	11 (9.2%)	14 (10.6%)	4 (3.7%)	10 (6.9%)	14 (10.1%)
		6. 動機と価値観	12 (10.0%)	6 (4.5%)	12 (11.1%)	14 (9.7%)	13 (9.4%)
		7. 必要性欠如	21 (17.5%)	10 (7.6%)	1 (0.9%)	12 (8.3%)	10 (7.2%)
	説明責任	8. 不確実性とリスク	5 (4.2%)	5 (3.8%)	2 (1.9%)	2 (1.4%)	3 (2.2%)
		9. 不公平性	3 (2.5%)	2 (1.5%)	1 (0.9%)	0 (0.0%)	0 (0.0%)

件では Diverse は Close より高く (A-3：24.5% vs. 19.4%，W-2：34.3% vs. 22.0%)，7 件では Diverse は Close より低かった (A-3：30.3% vs. 21.6%，W-2：29.7% vs.22.5%)．

代表的なコメントとして，管理可能性を個人の裁量に帰すもの (“時間や金銭の負担ではなく，自分の主体性次第で解決できる”，P928，Close-7) や，努力の程度を小さく見積

もるもの (“時間はかかるが，難なく実行できる”，P1341, Diverse-3；“それほどの努力は要らないと感じた”，P1398, Diverse-3) が確認された．

5.2.2 実現性・実用性

被験者は実現性の判断 (W-3) や難易度一般 (A-7) で達成可能性を評価し，“多くの人が審査に通らなくなる”（P304, Close-1）のように実現尤度 (R-1) において基準自体の尤もらしさを気にしていた．また，“収入を増やす”(P793, Diverse-3)，“持ち家を得る”(P1460，Close-3) など，具体的な利得 (W-4) も考慮した．選択肢数の増加にともない，W-3 は段階的に増え，A-7 は一度減った後に増え，R-1 は増えた後に頭打ちとなり，W-4 は 3 件で最大となった．一貫性欠如（A-6）の頻度は全条件で低水準にとどまった．

5.2.3 価値観・生活

選択肢が増えるにつれ，説明合理性は消極的受容 (R-3) から内省 (R-2) へと比重が移った．前者は “肩書で判断されるなら仕方ない”(P101，Close-7) のように諦めるもの，後者は “自分は大卒でも管理職でもない”(P904，Close-7) など自己に焦点を当てるものが多く確認された．実行可能性では心理的抵抗 (A-8) と実生活の影響 (A-9) が論じられ，後者は 3 件で最も顕著で，しばしば否定的であった（“生活様式を変えるのは難しい”，P2098，Close-3）．心理的抵抗は実務的な言い回しで表現されることが多く，“勤務時間を増やしたくない”(P1426，Close-3) などが典型例であった．

これに加え，実行意欲には動機と価値観 (W-6) と必要性欠如 (W-7) が確認された．前者には “東京に持ち家が欲しい”（P396，Diverse-3），後者には “ローンのために居住地を変える必要が本当にあるのか？”（P2529，Close-7）といった記述が含まれた．3 件では Diverse のほうが Close よりも心理的抵抗感（W-5）と必要性欠如（W-7）が低く（W-5: 3.7% vs. 10.6%，W-7: 0.9% vs. 7.6%），動機と価値観（W-6）は高かった（11.1% vs. 4.5%）．

5.2.4 説明責任

審査基準への批判である妥当性の指摘 (R-4) や明瞭性欠如 (R-5) は，選択肢数の増加とともに減少した．一方で，不公平性 (R-7)，AI 不信 (R-8)，AI 忌避 (A-11) などの対立的態度は，全体として低水準にとどまった．不確実性とリスクについては評価観点により傾向が異なり，説明合理性においては 3 件でピークを示した後に低下し (R-6)，実行意欲においては選択肢数の増加にともない減少した (W-8)．実行可能性では多様性基準の違いが一貫して観察され，不確実性とリスク (A-10) は 3 件と 7 件のいずれでも Diverse が Close より高かった（7.8% vs. 2.2%，8.2% vs. 1.4%）．

6. 考察

6.1 結果の解釈

3 件提示条件において Diverse 集合は Close 集合に比べて実行意欲を向上させた (5.1.1 節)．Diverse の被験者は，Close よりも有意義で（W-6：4.5% vs. 11.1%；W-7：3.8% vs. 1.9%），心理的抵抗感の少ない（A-8：9.7% vs. 5.9%；W-5：10.6% vs. 3.7%）プランを発見したことが意欲向上に結びついたと解釈できる．さらに，Diverse は Close よりも理解に要する労力を低減させた (5.1.2 節)．一般に，反事実的思考は認知的負荷を招きやすく [4], [5]，多様化はその負担を増やすと予想されたが，それとは逆の結果となった．定性的結果に見られたように，Diverse では Close よりも審査基準への批判が減り (R-5：8.8% vs. 5.4%)，管理可能性が高まった (A-3：19.4% vs. 24.5%) ことから，基準への疑義や計画の検討が減ったことで労力感が緩んだと考えられる．

3 件提示条件において Diverse 集合が持つこれらの利点は，7 件提示条件では非有意まで縮小した．集合の規模が大きい時、有望な反実仮想サンプルの候補が枯渇すると両集合において選択される反実仮想サンプルは互いに被り始める．本論文ではこれを候補飽和と呼ぶ．つまり，選択肢数の増加にともない両集合の性質差は縮退する傾向にある．これは，被験者がリコース集合に対して認識した多様性の評価値において，3 件提示条件よりも 7 件提示条件において両集合の差が小さかったこと (3.8 節) とも一貫する．

候補飽和により両条件の差は収束するにも関わらず，7 件条件において Diverse 集合は Close 集合よりも高い精神的要求を示した (5.1.2 節)．Diverse では集合が大きくなると内省が増え（R-2：17.2% vs. 24.0%），不確実性とリスクへの懸念が増し（A-10：7.8% vs. 8.2%），消極的受容が減った（R-3：7.5% vs. 3.9%）．これが起きた要因として，特徴量の変更方向の不整合性が挙げられる．L1 ノルムに基づく多様化では，反実仮想サンプル同士が異なる特徴を変える場合だけでなく，同じ特徴を反対方向に変える場合 (例：“勤務時間 +2 時間” を提案するリコースと “勤務時間 −2 時間” を提案するリコースを含む集合) でも効率的に相互距離を得る．後者を便宜的に符号衝突と呼ぶ．前者の方が数は多いため，リコース集合が小さい時は符号衝突は起こりにくい．しかし，リコース集合が大きい時に，候補飽和が進むにつれ相互距離維持のために Diverse は符号衝突を含みがちになる．各条件において符号衝突が提示された被験者の割合を調べたところ，これと一貫する結果が確認された (Close-3：16.4%，Diverse-3：11.1%；Close-7：29.2%，Diverse-7：35.8%)．つまり Diverse の被験者は，符号衝突によって基準への疑義や不確実性を強めたことで，認知的活動が要求されたと感じた可能性がある．これは，意思決定においては複数の目的が競合すると認知負荷が増すという知見 [15] と一貫する．

興味深いことに，多様化は選択肢数によらず否定的感情を抑える傾向を示した（5.1.2 節）．反事実的思考は後悔や罪悪感を誘発しうる [6] ため，予想に反する結果である．社会心理や行動変容の既存研究によれば，同一メッセージ

の反復は苛立ちや退屈を増やす一方 [7]，内容の多様化によってそれらは緩和される [16]．本実験においても，多様化が集合内の冗長性を低減したことが否定的感情の抑制に寄与した可能性がある．これを確認するため，集合内の全てのリコースが少なくとも 1 つの同一特徴量を変更したか否かを調べた．本論文ではこの指標を共通変更と呼ぶ．各条件において共通変更の割合を計算した結果，同一件数内においていずれも Diverse の方が低かった（Close-3：58.6% vs. Diverse-3：35.7%；Close-7：31.0% vs. Diverse-7：27.9%）．つまり，Diverse はリコース間で同一特徴の変更を避けたことで否定的感情を抑制したと解釈できる．

6.2 実応用

6.2.1 候補飽和に配慮した多様性の制御

Diverse 集合は 3 件条件において Close 集合よりも大きな心理的コストを強いることなく実行意欲を高めた．7 件条件では候補飽和によって両者の差は縮退し，むしろ符号衝突により Diverse 集合は高い認知負荷を被験者に課した．これらの結果を踏まえ，リコース集合が小さい場合は多様化の適用を基本とし，リコース集合が大きい場合は候補飽和に留意しながら適用判断を下すことを我々は推奨する．

運用時には，共通変更を通じて反実仮想サンプル間の冗長性を検知する，符号衝突に基づいて集合内の不整合性を確認する，といった対応により候補飽和の兆候を捉えられる．兆候が見られた場合には，リコース集合の大きさを小さく保つ，変更される特徴量数を明示的に制限する，候補となる反実仮想サンプルのデータプールを拡充するなどによって有効に対処できると考えられる．

6.2.2 多様化によるパーソナライゼーションの強化

個々のユーザの嗜好を対話的に捕捉し反映させたリコースを提示する推薦技術の研究が進められている [9], [33], [39]．しかし，高リスクな意思決定においてユーザはしばしば自身の要望を事前に明確化することができず [2]，不十分なデータに基づく個別化によって局所最適に陥る場合がある [19], [27]．この問題は，リコース集合の多様化を通じて広範な選択肢を提供することで対処できる可能性がある．

例えば，多様化されたリコース集合を与え，それに対するフィードバックを得て，リコースを個別化する (もしくはこれらを繰り返す) 方法が有効かもしれない．ユーザは受動的に与えられた幅広いオプションを評価できるため，局所最適を回避する個別化が可能となる．また，この多様化先行・個別化後行は，小さい集合では多様化が有利で，大きい集合では候補飽和で効用が縮退しうるという解釈と一貫する．インタラクションを繰り返す過程において，初期段階に広く嗜好を探索し，冗長性 (例：共通変更) の増加や不整合性 (例：符号衝突) の上昇といった候補飽和のシグナルが顕在化する前に個別化へ移行するのが望ましい．

6.3 本研究の限界と今後の課題

本研究は，リコースの多様化による効用には反実仮想サンプル間の相互差異性だけでなく，冗長性（例：共通変更），不整合（例：符号衝突），もしくは候補飽和が関与する可能性を示した．ただし，これらはあくまで多様化によって誘引されうる性質であり，本実験において実験条件として直接操作されてはいない．これらがユーザの反応に及ぼす因果効果を解明するためには，シミュレーション等の理論研究によって多様化の構造的な挙動を明らかにし，統制実験によってユーザの反応に与える影響を検証する必要がある．

また本研究は, 既存研究 [10], [12], [20], [31], [32], [36], [39] の動向を踏まえ，分野の中核テーマである与信をシナリオに用いた．このため，医療や法律など他の高リスク領域での再現性は今後検証される必要がある．

7. 結論

本研究はリコース集合の多様化による作用機序を解明するため，自動車ローンを題材とする統制実験 (N=750) において，実験的に多様性と選択肢数が操作されたリコース集合に対する被験者の反応を調べた．選択肢数が少ない時 (3 件)，多様化は有意義で心理的に抵抗感のない行動計画の発見を促し，被験者の実行意欲を高めた．反対に選択肢数が多い時 (7 件)，多様なリコース集合にはしばしば符号衝突が含まれ，被験者は審査基準や不確実性に対する疑念を感じ，高い認知労力を報告した．本研究は多様化による利益とコストのトレードオフ関係を実証的に分析したことで，理論と実践をつなぐ新たな知見を獲得した．さらに，実験結果と事後的観察に基づいて，冗長性 (共通変更) や不整合性 (符号衝突) を手がかりに候補飽和を回避することで多様化の効用を最大化する方法について論じた．

アルゴリズミック・リコースの根幹には，AI システムから不利な扱いを受けた個人を救済したいという思想がある．我々は本研究の知見が説明可能 AI 領域の発展に寄与し，最終的に助けを必要とする人へ届くことを切に願う．

参考文献

[1] Barocas, S., Selbst, A. D. and Raghavan, M.: The hidden assumptions behind counterfactual explanations and principal reasons, *Proc. of FAT**, pp. 80–89 (2020).
[2] Binns, R., Kleek, M. V., Veale, M., Lyngs, U., Zhao, J. and Shadbolt, N.: 'It's Reducing a Human Being to a Percentage'; Perceptions of Justice in Algorithmic Decisions, *Proc. of CHI*, pp. 1–14 (2018).
[3] Bove, C., Lesot, M. J., Tijus, C. A. and Detyniecki, M.: Investigating the Intelligibility of Plural Counterfactual Examples for Non-Expert Users: an Explanation User Interface Proposition and User Study, *Proc. of IUI*, pp. 188–203 (2023).
[4] Byrne, R. M. J.: The Rational Imagination: How People Create Alternatives to Reality, *The Rational Imagination* (2005).
[5] Byrne, R. M.: Précis of the rational imagination: How

people create alternatives to reality, *Behavioral and Brain Sciences*, Vol. 30 (2007).

[6] Byrne, R. M.: Counterfactual Thought, *Annual Review of Psychology*, Vol. 67, pp. 135–157 (2016).

[7] Cacioppo, J. T. and Petty, R. E.: Effects of message repetition and position on cognitive response, recall, and persuasion., *Journal of Personality and Social Psychology*, Vol. 37, pp. 97–109 (1979).

[8] Cohen, J.: *Educational and Psychological Measurement*, Vol. 20, pp. 37–46 (1960).

[9] Esfahani, S., Toni, G. D., Lepri, B., Passerini, A., Tentori, K. and Zancanaro, M.: Preference Elicitation in Interactive and User-centered Algorithmic Recourse: An Initial Exploration, *Proc. of UMAP*, pp. 249–254 (2024).

[10] Gemalmaz, M. A. and Yin, M.: Understanding Decision Subjects' Fairness Perceptions and Retention in Repeated Interactions with AI-Based Decision Systems, *Proc. of AIES*, pp. 295–306 (2022).

[11] Hart, S. G. and Staveland, L. E.: *Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research*, Vol. 52, pp. 139–183 (1988).

[12] Karimi, A. H., Barthe, G., Schölkopf, B. and Valera, I.: A Survey of Algorithmic Recourse: Contrastive Explanations and Consequential Recommendations, *ACM Computing Surveys*, Vol. 55 (2022).

[13] Karimi, A.-H., Schölkopf, B. and Valera, I.: Algorithmic Recourse: from Counterfactual Explanations to Interventions, *Proc. of FAccT*, pp. 353–362 (2021).

[14] Kirfel, L. and Liefgreen, A.: What If (and How ...)? - Actionability Shapes People's Perceptions of Counterfactual Explanations in Automated Decision-Making, *ICML-21 Workshop on Algorithmic Recourse* (2021).

[15] Kivikangas, J. M., Vilkkumaa, E., Blank, J., Harjunen, V., Malo, P., Deb, K., Ravaja, N. J. and Wallenius, J.: Effects of many conflicting objectives on decision-makers' cognitive burden and decision consistency, *European Journal of Operational Research*, Vol. 322, pp. 182–197 (2025).

[16] Kocielnik, R. and Hsieh, G.: Send Me a Different Message: Utilizing Cognitive Space to Create Engaging Message Triggers, *Proc. of CSCW*, pp. 2193–2207 (2017).

[17] Landis, J. R. and Koch, G. G.: The Measurement of Observer Agreement for Categorical Data, *Biometrics*, Vol. 33, p. 159 (1977).

[18] Laugel, T., Jeyasothy, A., Lesot, M.-J., Marsala, C. and Detyniecki, M.: Achieving Diversity in Counterfactual Explanations: a Review and Discussion, *Proc. of FAccT*, pp. 1859–1869 (2023).

[19] Li, L., Chu, W., Langford, J. and Schapire, R. E.: A Contextual-Bandit Approach to Personalized News Article Recommendation, *Proc. of WWW*, pp. 661–670 (2010).

[20] Lyons, H., Wijenayake, S., Miller, T. and Velloso, E.: What's the Appeal? Perceptions of Review Processes for Algorithmic Decisions, *Proc. of CHI*, pp. 1–15 (2022).

[21] McEleney, A. and Byrne, R. M. J.: Spontaneous counterfactual thoughts and causal explanations, *Thinking & Reasoning*, Vol. 12, pp. 235–255 (2006).

[22] Medin, D. L., Goldstone, R. L. and Gentner, D.: Respects for similarity., *Psychological Review*, Vol. 100, pp. 254–278 (1993).

[23] Miller, G. A.: The magical number seven, plus or minus two: Some limits on our capacity for processing information., *Psychological Review*, Vol. 63, pp. 81–97 (1956).

[24] Miller, T.: Explanation in artificial intelligence: Insights from the social sciences, *Artificial Intelligence*, Vol. 267, pp. 1–38 (2019).

[25] Mothilal, R. K., Sharma, A. and Tan, C.: Explaining machine learning classifiers through diverse counterfactual explanations, *Proc. of FAT**, pp. 607–617 (2020).

[26] Nicolle, A., Bach, D. R., Frith, C. and Dolan, R. J.: Amygdala involvement in self-blame regret, *Social Neuroscience*, Vol. 6, pp. 178–189 (2011).

[27] Rashid, A. M., Albert, I., Cosley, D., Lam, S. K., McNee, S. M., Konstan, J. A. and Riedl, J.: Getting to Know You: Learning New User Preferences in Recommender Systems, *Proc. of IUI*, pp. 127–134 (2002).

[28] Ryan, R. M. and Deci, E. L.: Self-determination theory and the facilitation of intrinsic motivation, social development, and well-being., *American Psychologist*, Vol. 55, pp. 68–78 (2000).

[29] Scheibehenne, B., Greifeneder, R. and Todd, P. M.: Can There Ever Be Too Many Options? A Meta-Analytic Review of Choice Overload, *Journal of Consumer Research*, Vol. 37, pp. 409–425 (2010).

[30] Spiro, R. J., Feltovich, P. J., Coulson, R. L. and Anderson, D. K.: *Multiple analogies for complex concepts: antidotes for analogy-induced misconception in advanced knowledge acquisition*, pp. 498–531 (1989).

[31] Tominaga, T., Yamashita, N. and Kurashima, T.: Reassessing Evaluation Functions in Algorithmic Recourse: An Empirical Study from a Human-Centered Perspective, *Proc. of IJCAI*, pp. 7913–7921 (2024).

[32] Tominaga, T., Yamashita, N. and Kurashima, T.: The Role of Initial Acceptance Attitudes Toward AI Decisions in Algorithmic Recourse, *Proc. of CHI*, pp. 1–20 (2025).

[33] Toni, G. D., Viappiani, P., Teso, S., Lepri, B. and Passerini, A.: Personalized Algorithmic Recourse with Preference Elicitation, *Transactions on Machine Learning Research* (2024).

[34] Ustun, B., Spangher, A. and Liu, Y.: Actionable Recourse in Linear Classification, *Proc. of FAT**, pp. 10–19 (2019).

[35] Venkatasubramanian, S. and Alfano, M.: The philosophical basis of algorithmic recourse, *Proc. of FAT**, pp. 284–293 (2020).

[36] Verma, S., Boonsanong, V., Hoang, M., Hines, K., Dickerson, J. and Shah, C.: Counterfactual Explanations and Algorithmic Recourses for Machine Learning: A Review, *ACM Computing Surveys*, Vol. 56, pp. 1–42 (2024).

[37] Wachter, S., Mittelstadt, B. and Russel, C.: Counterfactual Explanations Without Opening the Black Box: Automated Decisions and thhe GDPR, *Harvard Journal of Law & Technology*, Vol. 31, pp. 842–887 (2018).

[38] Wang, D., Yang, Q., Abdul, A. and Lim, B. Y.: Designing Theory-Driven User-Centric Explainable AI, *Proc. of CHI*, pp. 1–15 (2019).

[39] Wang, Z. J., Vaughan, J. W., Caruana, R. and Chau, D. H.: GAM Coach: Towards Interactive and User-centered Algorithmic Recourse, *Proc. of CHI* (2023).

[40] Zeelenberg, M. and Pieters, R.: A Theory of Regret Regulation 1.0, *Journal of Consumer Psychology*, Vol. 17, pp. 3–18 (2007).