

オンライン議論における公正性の比較分析: ファシリテーションの違いに注目して

辰巳 智行¹ 中澤 高師¹ 福田 直樹¹ 吉田 寛¹ 青木 美奈¹ 大石 哲也¹ 加藤 史也¹ 中島 大晴¹
山口 七海¹

概要: 本稿は、ファシリテーション手法の違いが議論にどのような影響を与えるかを公正性に注目して明らかにする。「高齢者の運転免許返納問題」をテーマに、大学生 117 人を無作為に 6 つのグループに振り分け、グループごとにオンライン議論実験を実施した。グループは、ファシリテーションを行わない 2 群、人間がファシリテーションを行う 2 群、IBIS 構造を利用した自動ファシリテーションを行う 2 群であり、議論の前後にアンケート調査を実施した。主観的不公正の評価を被説明変数とした重回帰分析の結果、ファシリテーションの存在が不公正感を強めていることが分かった。そのなかでも「他の参加者の発言の中で、特定の意見や立場を不当に強く否定する意見があった」という評価を、特に自動ファシリテーションが強めている可能性があることが明らかになった。

A Comparative Analysis of Fairness in Online Discussions: Examining the Effects of Different Facilitation Methods

1. はじめに

本研究の目的は、議論支援システムを用いたオンライン議論実験によって、ファシリテーションが参加者の主観的な公正性評価にどのような影響を与えるのかについて明らかにすることである。

オンライン上での議論は時間や空間による制約を乗り越える可能性がある一方で、無秩序な発言により議論が成立しなかったり、意見の精鋭化によって集団分極化が生じることが危惧されている [1]。そのため、議事進行や情報整理を専門とするファシリテータの役割に注目が集まっている。

議論において、ファシリテータ自身は原則として中立な立場で、意思決定にも直接関わることはない。しかし、ファシリテータのパーソナリティ（ジェンダー、世代、人種など）や、議事進行中の発言や非言語的な挙動が議論を左右してしまうことに対する危惧もあり、特に政治的に対立する場面や科学的な調査ではファシリテーションの標準化が望まれている。例えば、討論型世論調査の手法を開発した Fishkin らは、ファシリテータ（彼らはモデレーターと

呼ぶ）に対して、参加者に中立的で公平な発言機会を提供するだけでなく、参加者の発言を要約しないことや「正しい」知識を提供しないことを求め、事前研修のトレーニングを通じてファシリテータ間の差異をなくすことを強調する [2]。そのため、オンライン議論のように 24 時間継続的に投稿がなされる議論空間では、十分にトレーニングされたファシリテータを常に配置することは難しく、また、投稿数が爆発的に増加すると人力による対応は困難となる [3]。議論空間のスケールアップのために、人工知能技術を用いた自動ファシリテーションの開発が行われている [4][5]。

しかし、人工知能技術による意思決定支援では、アルゴリズムや学習データの潜在的バイアスによって意図しない誘導が起きたり、悪意のあるユーザがアルゴリズムの間隙をつくことで議論の公平性が歪められる可能性が考えられる。また、アルゴリズムやテンプレートの不備で、人間のファシリテーションではあり得ない行動や発言を行うことで、参加者がシステムや主宰者に不信をいだき、信頼関係が破綻する恐れもあることから、人工知能技術によるファシリテーションが議論に与える影響について検証する必要がある。

本研究では、人工知能技術によるファシリテーションが

¹ 静岡大学情報学部
Faculty of Informatics, Shizuoka University, 3-5-1, Johoku,
Naka, Hamamatsu, Shizuoka 432-8011, Japan

議論の公正性に与える影響を明らかにする。議論を評価する上で、公正性は最も重要な価値の一つである。一般的に公正性は多義的な概念であり、様々な解釈がありうる [6][7]。議論における公正性にも様々な類型がありうる。また、発言回数など客観的な指標の他に、自分の意見が尊重されていると感じるかどうか等、主観的にも捉えることができる。本稿では、「高齢者の自動車免許返納問題」をテーマとした議論実験によって、人工知能技術による自動ファシリテーションが議論参加者の主観的な公正性の感覚に与える影響を検証した。

2. 方法

2.1 実験参加者

実験は、2020年1月に静岡県内の大学で、教養系科目を受講する18歳から22歳までの117名の学部生（情報系および工学系）を対象とした。参加者を無作為に6つのグループに分け、実験時間中は他グループの議論を参照できない状況にした。参加者は匿名化され、参加者間も実験主宰者も議論中に発言した個人が特定できない状態として、参加者にも事前に説明した。ひとつのグループの参加者数は20人程度を目安とした。これは議論が停滞しにくい最低限の参加者人数として、事前の複数回の予備実験の結果から導き出したものである。

2.2 実験システム

実験では、オンライン議論支援システムとしてD-Agreeを用いた。D-Agreeは名古屋工業大学の伊藤孝行研究室を中心とした研究プロジェクトが開発したオンライン上の非同期型議論支援・意見集約システムで、文字テキストをベースとしたスレッド型議論フォーラムである [8]。システムの特徴として、Issue-Based Information System (IBIS) 構造を利用した自動ファシリテーション機能を持つ。この機能は、参加者の投稿を、イシュー（課題）、アイデア、およびアイデアに対する賛否の4つに自動分類する。そして、その分類に応じて参加者に質問を自動的に尋ねて、参加者の投稿を促す仕組みになっている。自動ファシリテーションは定型文を用いて行われ、(1) 投稿された課題についてどう思うかアイデアを求める発言、(2) 課題に対するアイデアのメリットやデメリットはなにかを尋ねる発言、(3) 課題の解決策はどのようなものかを尋ねる質問、(4) さらなるアイデアの追加を促す発言の四種類に分類される。ただし、実験参加者は、このIBIS構造を意識する必要はなく、議論時間内であれば自由に発言することが許される。また、ファシリテーションの問いかけに対して必ずしも応じなくともよい。

議論実験は、IBIS構造を利用した自動ファシリテーションを行う2群を実験群とし、対照群として人間がファシリ

テーションを行う2群、ファシリテーションを行う主体がない2群を設定した（合計6群）。人間がファシリテーションを行う群では、それぞれ1名のファシリテータが、自動ファシリテーションに用いられるのと同じ定型文のみを投稿する形で議論した。

参加者は、まず議論前アンケートに回答し、全員の回答が終わった段階でグループごとに約50分間の議論に参加した。そして、議論終了時に「免許返納制度に関する意見」のみを尋ねて（中間アンケート）、自分が議論したグループ内の意見分布を周知した。最後に、議論や中間アンケートの結果を踏まえた上で、議論後アンケートに回答した。中間アンケートを実施したのは、参加者が参加したグループの意見分布を知ること、自身が議論空間で多数派に属していたのか、それとも少数派であったのかを認識した上で、議論やファシリテーションの評価を回答することを求めたためである。アンケート集計はGoogleフォームを用いた。

2.3 議論テーマ

議論テーマは「高齢者の自動車免許返納問題」とした。高齢者の運転は加齢に伴う身体機能の変化や認知機能の低下から事故などのリスクが高いと指摘されている。一方で、日本の地方部においては、高齢者が自由に移動できる交通手段として自動車に依存しているという事情がある。現在の制度では、70歳以上の免許更新時に認知機能検査や高齢者講習受講が義務化されているものの、65歳以上免許保持者は1863万人以上、保有者全体の約22%を占める [9]。しかし、近年、高齢者による危険運転や事故が目目される中で、高齢者の一律免許返納の制度化にも議論が及んでいる。注目度が高く、前提知識がなくても議論でき、参加者自身やその家族の問題として身近に感じられるテーマであることから採用した。

2.4 データ

実験によって、議論ログとアンケート結果を得た。実験参加者117名のうちオンライン議論で発言したのは115名、3回のアンケート全てと議論に参加した参加者は107名であった（アンケートの有効回答は議論前:110、議論中間:109、議論後:112）。議論内の参加者の行動に関しては表1を参照されたい。

アンケートは(1) 免許返納に関する意見や高齢者の運転に対する態度に関する設問、(2) 参加者自身の運転経験や家族の運転状況など当事者性に関わる設問、(3) 議論内容や議論の進行に関する参加者の評価に関する設問、および(4) 参加者の学部、性別や年齢などの基本属性で構成される。このうち、議論前アンケートでは(1)と(2)と(4)を、中間アンケートでは(1)のうち免許返納制度に関する意見のみを、議論後アンケートでは(1)と(3)を尋

表 1 議論実験参加者の行動
 Table 1 Participant's Behavior in the Discussion.

	G1	G2	G3	G4	G5	G6	計
ファシリテーション	AI	AI	人間	人間	なし	なし	
実験参加者 (人)	20	21	20	19	19	19	117
議論発言者 (人)	18	18	20	19	19	19	113
参加者発言数 (回)	64	79	70	57	64	97	431
参加者 1 人あたりの発言数 (平均± SD)	3.6 ± 4.6	4.4 ± 4.7	3.5 ± 3.7	3.0 ± 2.4	3.4 ± 3.0	5.1 ± 9.3	3.8 ± 5.2
発言あたりの文字数 (中央値)	61.5	161	117	168	82	121	116
ファシリテータ発言数	21	26	13	15	-	-	-

ね、議論による意見変容を測定できるようにした。

2.5 分析手法

分析は、不公正性指数を被説明変数とした重回帰分析を行った。被説明変数と説明変数は下記の項目である。

不公正性指数 不公正性指数は、参加者が議論を通じて感じた主観的な不公正感覚を測定するもので、議論の中での自らの立ち位置や議論進行への評価など7項目で構成される（「自分の発言は他の参加者から尊重されていた（逆転項目）」「自分の言いたい意見を発言出来ないことがあった」「自分はほかの参加者よりも知識量で劣っているように感じた」「自分の立場は議論の中では少数派であると感じた」「他の参加者の発言の中で、特定の意見や立場を不当に強く否定する意見があった」「他の参加者の発言に趣旨の理解できない意見があった」「過剰に持ち上げられている意見や立場があった」）。それぞれの回答項目は「そう思う」から「そう思わない」までの5件法で尋ねて、「そう思う」場合を5とし「そう思わない」に向かい4から1を与え、ネガティブな印象で得点が高くなるようにした（逆転項目はそれと逆に点数を与えた）。不公正性指数はこの項目の合計得点である。

ファシリテーション ファシリテーションの方法に応じて、グループごとにダミー変数に変換した。ファシリテーションなしのグループ参加者を0として、人間ファシリテータがいたグループ、人工知能技術による自動ファシリテータ (AI ファシリテータ) がいたグループの参加者それぞれを1とした。

免許制度意見ダミー 「あなたは、75歳以上の免許の返納制度について、どう考えていますか」という設問について「返納制度は必要ない」という回答を0として、「運転者が自主的に判断して返納する制度にすべき（自主返納）」「運転能力や必要性の審査に通過した高齢者のみが運転できる制度にすべき（能力審査）」「年齢で一律に返納させる制度にすべき（一律返納）」の3つのダミー変数を生成した。

関心ダミー 「高齢者の運転事故にこれまで関心がありま

したか?」という項目で、「知ってはいたが、関心はなかった」を0、「関心がある」を1としてダミー変数とした。なお「(高齢者の事故を) 知らなかった」という回答はなかった。

当事性ダミー 高齢者事故と参加者の関わりについて「あなたや身近な人が、高齢者の運転で事故に遭ったり、被害を受けそうになったりした経験はありますか?」という設問で「ない」という回答を0、「ある」を1としてダミー変数を生成した。

性別ダミー 女性を0として、男性を1とした。

他者意見尊重指数 参加者が議論で発言する際に、どの程度他者の意見や考え方を尊重していたのかについて測定したもので、次の3項目から構成される。「自分は意見を投稿する前に、他の参加者の意見を把握しようとした」「自分は他の参加者の発言を尊重しようとした」「自分は免許返納の影響を受ける当事者の立場を考慮して意見を出した」。それぞれの回答項目は「そう思う」から「そう思わない」までの5件法で、「そう思う」場合を5とし「そう思わない」に向かい4から1を与え、尊重していたという自己認識が高いほど得点が高くなるようにした。他者意見尊重指数はこれら3項目の合計得点である。

議論アクティビティ オンライン議論における参加者の活動を示すものである。発言回数は参加者の投稿回数で、返信数は発言回数のうち他の投稿に対して返信した数である。被返信回数は自分の返信に対して他の参加者やファシリテータが返信をした数である。発言文字数は参加者が実験中に投稿した文字数の合計で、発言平均文字数は発言文字数を発言数で除した数である。ポイントは、オンライン議論支援システムが議論への貢献や他の参加者からの賛同をもとに参加者に付与する得点である [10]。

3. 結果と考察

不公正性指数が (1) 参加者の議論前からの意見に影響を受けるのか、それとも (2) 議論後の意見に影響を受けるのか、あるいは (3) 意見の変化によって生じる可能性が

表 2 主観的不公正指標を被説明変数とした重回帰分析の結果

Table 2 Results of Multiple Regression Analysis of the Subjective Evaluation of Unfairness

説明変数 (括弧内は範囲)	モデル 1			モデル 2			モデル 3		
	β	p -value		β	p -value		β	p -value	
ファシリテータダミー									
人間 (0,1)	2.11	0.06	†	2.11	0.05	†	1.99	0.07	†
AI(0,1)	2.52	0.04	*	2.69	0.02	*	2.54	0.03	*
免許制度意見ダミー (ref=必要なし)									
自主返納-議論前 (0,1)	-2.60	0.59							
能力審査-議論前 (0,1)	-2.28	0.62							
一律返納-議論前 (0,1)	-2.96	0.53							
自主返納-議論後 (0,1)				-1.20	0.73				
能力審査-議論後 (0,1)				-3.77	0.25				
一律返納-議論後 (0,1)				-3.79	0.28				
免許制度意見変化 (ref=要件緩和)									
変化なし (0,1)							-2.03	0.10	
厳格化 (0,1)							-2.79	0.09	
関心ダミー (0,1)	-1.02	0.29		-0.85	0.37		-1.04	0.26	
当事者性ダミー (0,1)	0.63	0.56		0.26	0.81		0.17	0.87	
性別ダミー (0,1)	-1.44	0.34		-1.70	0.24		-2.02	0.17	
他者意見尊重指数 (3,15)	-0.22	0.40		-0.21	0.41		-0.17	0.48	
発言回数 (0,43)	0.06	0.81		-0.01	0.97		-0.01	0.98	
返信回数 (0,17)	0.38	0.31		0.40	0.28		0.43	0.24	
被返信回数 (0,22)	-0.30	0.28		-0.28	0.30		-0.28	0.31	
発言文字数 (0,1352)	0.00	0.66		0.00	0.65		0.00	0.76	
発言平均文字数 (0,344)	0.00	0.81		-0.00	0.94		0.00	0.93	
ポイント (0,823)	-0.01	0.29		-0.01	0.36		-0.01	0.38	
切片	24.80	<0.001	***	25.69	<0.001		24.24	<0.001	
決定係数		0.14			0.18			0.16	
調整済み決定回数		-0.01			0.04			0.04	
モデル適合度 (p -value)		0.56			0.22			0.23	

N = 107, Significant Codes: ***: <0.001, **: <0.01, *: <0.05, †: <0.10

表 3 「強い否定を意見」の認知を被説明変数とした重回帰分析の結果

Table 3 Results of Multiple Regression Analysis of the Perceived Negative Reactions as Unfairness

説明変数 (括弧内は範囲)	モデル 1			モデル 2			モデル 3		
	β	p -value		β	p -value		β	p -value	
ファシリテータダミー									
人間 (0,1)	1.02	<0.001	***	0.98	<0.001	***	1.02	<0.001	***
AI(0,1)	1.44	<0.001	***	1.43	<0.001	***	1.41	<0.001	***
免許制度意見ダミー (ref=必要なし)									
自主返納-議論前 (0,1)	0.01	1.00							
能力審査-議論前 (0,1)	0.23	0.85							
一律返納-議論前 (0,1)	-0.13	0.92							
自主返納-議論後 (0,1)				0.89	0.33				
能力審査-議論後 (0,1)				0.60	0.48				
一律返納-議論後 (0,1)				0.02	0.98				
免許制度意見変化 (ref=要件緩和)									
変化なし (0,1)							-0.18	0.57	
厳格化 (0,1)							-0.54	0.21	
関心ダミー (0,1)	-0.09	0.73		-0.08	0.76		-0.10	0.67	
当事者性ダミー (0,1)	0.10	0.73		-0.06	0.84		0.04	0.89	
性別ダミー (0,1)	-0.17	0.67		-0.24	0.53		-0.30	0.44	
他者意見尊重指数 (3,15)	0.04	0.50		0.05	0.49		0.07	0.31	
発言回数 (0,43)	0.07	0.25		0.06	0.33		0.07	0.27	
返信回数 (0,17)	-0.02	0.81		-0.02	0.85		-0.02	0.87	
被返信回数 (0,22)	-0.01	0.93		-0.01	0.88		0.00	0.96	
発言文字数 (0,1352)	0.00	0.96		0.00	0.95		0.00	0.80	
発言平均文字数 (0,344)	0.00	0.45		0.00	0.44		0.00	0.34	
ポイント (0,823)	0.00	0.37		0.00	0.42		0.00	0.41	
切片	1.28	0.37		0.96	0.44		1.48	0.10	
決定係数		0.25			0.27			0.25	
調整済み決定回数		0.12			0.14			0.14	
モデル適合度 (p -value)		0.02			0.01			0.01	

N = 107, Significant Codes: ***: <0.001, **: <0.01, *: <0.05, †: <0.10

あるのかを確認するために、3つのモデルに分けて検討した。なお、分析では3つ全てのアンケートに回答した107名を分析対象としている。

不公正性指数全体の特徴について見ると、人間とAIの両方のファシリテータダムーが有意に影響しており、いずれも正の方向に作用している。これはファシリテータが参加しなかったグループよりもファシリテータが参加したグループで不公正性の感覚が強くなっていることを示している。さらに、AIファシリテータの方が人間によるファシリテータよりもその影響が強い。

そこで、不公正性指標を構成する7つの質問項目それぞれについて、質問項目の得点を被説明変数とし、その他の変数を説明変数とする重回帰分析を行ったところ、「他の参加者の発言の中で、特定の意見や立場を不当に強く否定する意見があった」という設問のみが、ファシリテータダムーの影響を受けていたことがわかった(表3)。ここからファシリテータの議論への介入、とりわけ自動ファシリテータによるそれが「特定の意見や立場を不当に強く否定する意見があった」という感覚を強化していると考えられる。

一方で、事後アンケートにおける議論内でのファシリテータへの評価を問う設問項目では、「議論においてファシリテータは必要であった」という設問と「議論の進行にファシリテータは影響を持っていた」という設問項目の回答の間には相関係数0.83という強い正の相関がみられた。即ちファシリテータの影響を強く感じた参加者はファシリテータの存在を肯定的に捉えている傾向が認められ、ファシリテータが介入し議論に影響することで議論の不公正感覚を高めるといった結果と一見矛盾する結果となった。

解釈のひとつとして考えられるのは、ファシリテーションが行われることで他の参加者の意見に言及して賛否を問うような発言が増え、その結果として特定の意見や立場を否定する言動がファシリテータの参加しなかったグループに比べ増加した可能性である。本実験で用いたIBIS構造に基づく自動ファシリテーションでは、アイデアについての意見や課題の指摘を促す定型文が多く含まれており、実際に議論で多く投稿されている。つまり、ある意見に対する賛否の確認や課題の指摘を促すファシリテーションが議論に影響を与え、ファシリテータの存在が肯定的に捉えられる一方で、特定の立場を強く批判する意見の出現にもつながった可能性がある。このように解釈すれば、ファシリテータの参加が「特定の意見や立場を不当に強く否定する意見があった」という不公正感に繋がったことと、ファシリテータの影響が肯定的に捉えられていることの両方が説明できる。

4. 結論

本研究は、人工知能技術による自動ファシリテーションが議論に与える影響について、主観的な公正性に着目して明らかにした。議論実験の結果から、ファシリテーションの存在が不公正感を強めていることが分かった。そのなかでも「他の参加者の発言の中で、特定の意見や立場を不当に強く否定する意見があった」という評価を、特に自動ファシリテーションが強めている可能性があることが明らかになった。

本実験の結果はファシリテータが議論に及ぼす影響の二面性を示唆している。IBIS構造に基づきアイデアへの賛否を含む意見を促すファシリテーションは、議論を深めるためには必要不可欠である。しかし、同時に、参加者がこうした意見を促すファシリテーションを「ファシリテータがアイデアを否定的に捉えている」と認知してしまうと、参加者の主観的な不公正感を増加させる可能性がある。

ただし、アイデアへの意見を促すファシリテーションがどのように議論内容に影響しているのかを明らかにするためには、コンテンツ分析を含む詳細な分析が必要になる。また、人間ファシリテーションよりも自動ファシリテーションの方が「他の参加者の発言の中で、特定の意見や立場を不当に強く否定する意見があった」という評価を強める効果があった理由については、本稿では扱うことが出来なかった。

人工知能技術による自動ファシリテーションの発展は、オンライン議論に新たな可能性をもたらしつつある。同時に、そのリスクについても慎重に検討する必要がある。主観的な公正性に着目して自動ファシリテーションの影響を検討した本稿は、そのための一助となるものである。

謝辞 本稿は、科学技術振興機構による戦略的創造研究推進事業『エージェント技術に基づく大規模合意形成支援システムの創成』(代表:伊藤孝行教授)(課題番号:JPMJCR15E1)の助成を受けたものである。

本実験にあたり静岡大学情報学部先端情報学実習「ICTを利用した新しい合意形成」受講学生の協力を得た。

参考文献

- [1] Sunstein, C. R.: *#Republic—Divided Democracy in the Age of Social Media*, Princeton University Press (2017).
- [2] エネルギー・環境の選択肢に関する討論型世論調査実行委員会: エネルギー・環境の選択肢に関する討論型世論調査調査報告書 (2012).
- [3] Fisher, M.: Inside Facebook's Secret Rulebook for Global Political Speech, *The New York Times*, (online), available from (<https://nyti.ms/2GIG2sn>) (2018).
- [4] 福田直樹, 福島俊一, 伊藤孝行, 谷口忠大, 横尾 真: 複雑化社会における意思決定・合意形成のためのAI技術, 人工知能, Vol. 34, No. 6, pp. 863–868 (オンライン), 入手先 (<http://id.nii.ac.jp/1004/00010382/>) (2019).

- [5] Fishkin, J., Garg, N., Gelauff, L., Goel, A., Munagala, K., Siu, A. and Yandamuri, S.: Deliberative Democracy with the Online Deliberation Platform, *the 7th AAAI Conference on Human Computation and Crowdsourcing (HCOMP 2019)*, Demonstration paper, Skamania Lodge, WA, USA, (online), available from <https://www.humancomputation.com/2019/assets/papers/144.pdf> (2019).
- [6] Deutsch, M.: Equity, Equality, and Need—What Determines which Value Will Be Used as the Basis of Distributive Justice?, *Journal of Social Issues*, Vol. 31, No. 3, pp. 137–149 (online), DOI: 10.1111/j.1540-4560.1975.tb01000.x (1975).
- [7] Stone, D. A.: *Policy Paradox—The Art of Political Decision Making*, W. W. Norton & Company, revised edition (2001).
- [8] Shibata, D., Moustafa, A., Ito, T. and Suzuki, S.: On Facilitating Large-Scale Online Discussions, *PRICAI 2019—Trends in Artificial Intelligence* (Nayak, A. C. and Sharma, A., eds.), Lecture Notes in Computer Science, Cham, Springer International Publishing, pp. 608–620 (online), DOI: 10.1007/978-3-030-29911-8_47 (2019).
- [9] 警察庁交通局運転免許課: 運転免許統計平成 30 年版 (2019).
- [10] Ito, T. [Takayuki], Imi, Y., Sato, M., Ito, T. [Takahiro] and Hideshima, E.: Incentive Mechanism for Managing Large-Scale Internet-Based Discussions on COLLAGREE, *Proceedings of the 3rd Collective Intelligence*, Vol. 2015, pp. 1–4 (2015).