

インフラ化するプラットフォームの「信頼」とは

武蔵大学准教授

宇田川 敦史

ソーシャルメディアの倫理を担う重要な要素のひとつに、情報の選別・分配を遂行するアルゴリズムのデザイン（設計）が挙げられる。このアルゴリズムはしばしば「ブラックボックス」とであると表象され、プラットフォーム企業がユーザーを「操作」するための「秘密」とであると批判的に論じられることも多い。このようなアルゴリズムのデザインは、確かに企業の利益の源泉にもなっているが、複雑性を縮減したいというユーザーの社会的な要請に適応的であるという側面も見逃せない。

ブルーノ・ラトゥールは、科学的な知識や技術一般における「ブラックボックス化 (blackboxing)」について以下のように定義している。「科学のおよび技術的な成果が、その成功自体によって不可視化されること。科学的な事実が確立している限り、もしくは機械が効率的に動作している限り、人々はそのインプットとアウトプットのみに注目することになり、内的な複雑さは無視されることになる。¹⁾」。

現代のメディア環境において、スマートフォンのみならずインターネット上のソーシャルメディア、プラットフォーム全般も、ラトゥールのいうブラックボックスになっている。ほとんどの場合、アテンションが向くのはそれらのアウトプットの一部だけであり、システムやアルゴリズムの内的な複雑さについてアテンションを向ける機会がないといえるからだ。しかし、このようなブラックボックス化は所与のものではなく、歴史的・社会的に構築されるものである（ラトゥールの議論はまさにそのプロセスの重要性を主張するものだ）。

ソーシャルメディアのアルゴリズムが、ブラックボックスとして社会的に扱われているということにも、一定の社会的なニーズがある。すなわち、アテンション・エコノミーという認知資源が欠乏した環境においてユーザーは、ソーシャルメディアの技術的な仕組みやアルゴリズムについて、いちいちアテンションを向けて吟味する余裕はない。流れていくコンテンツを追いかけるので精一杯であり、メディアの技術的な内部をブラックボックスとして扱うほうがはるかに効率的で、現実的なのである。

実際、Xは2023年に、おすすめタイムラインに表示されるポストがどのような基準で選別・配置されているかのアルゴリズムについて、（全部ではないが）公開をしている²⁾。しかしこのようにXのアルゴリズムが公開されているという事実を、どの程度のユーザーが認識しているだろうか。ソーシャルメディアが倫理的といえる振る舞いをしているのかどうかは、まさにそのアルゴリズムの設計がどのようになされているかと表裏一体のはずにもかかわらず、である。

この公開されたアルゴリズムを活用しようとするのは、むしろマーケティングを行う送り手の側である。ソーシャルメディアを利用して、ユーザーのアテンションをどのように集めるのか、まさにアテンション・エコノミーの論理のなかにおいて、いかに自社の情報を有利に表示させるのか、そのための「戦略」として、Xのおすすめアルゴリズムは「ハック」されてしまう。たとえば2025年7月に「X アルゴ

リズム」でウェブページをサーチすると、「拡散されやすいX（旧 Twitter）のアルゴリズムまとめ」「X (Twitter)のアルゴリズムとは？最新の傾向から運用のコツまで解説！」「X (Twitter)のアルゴリズムとは？おすすめタイムラインにポストを表示させる秘訣をご紹介！」などと、上位十件に表示されるものはすべて、Xのアルゴリズムを活用して拡散や上位表示をさせたい送り手に向けた解説記事になっている³。受け手であるユーザーは、ソーシャルメディアのタイムラインに何が表示されるのかをブラックボックスとして扱っているが、そのインターフェイスは、このような解説を手がかりに、送り手が「工夫」をした結果として構築されたものである。Xが「秘密主義批判」への応答としてアルゴリズムを公開したことはむしろ、送り手と受け手の間での非対称性を拡大し、社会的構築物としてのブラックボックスの機能を強める結果となっていることは皮肉ともいえるだろう。

このように、アルゴリズムのブラックボックスをめぐる非対称性の拡大は、プラットフォーム側の秘密主義によるものではない。認知資源が不足しているユーザー側の過剰ともいえる「信頼」が、逆説的に機能していると解釈すべきだろう。それでは、この「信頼（Trust）」とは何か、少し掘り下げて考えてみたい。

科学技術社会論において「インフラ（Infrastructure）」とは、それが安定的に稼働している時には「透明(transparency)」なものとして扱われるような事物の状態を指す⁴。アルゴリズムがブラックボックスとして社会的に扱えるということは、まさにそれが社会の「インフラ」として、その介在が背景化し無色透明かのようにみなされる状態にあることを示している。そして、それを可能にする社会的な態度こそが「信頼」であるといえるだろう。

たとえば、水道水の供給システムについて、私たちは普段、水の安全性がどのように確保されているのか、取水した水をどのように洗浄しており、水道管のネットワークはどのようになっているのか、などの仕組みの介在を意識することはほとんどない。これはまさに水道が社会のインフラとして扱われていることを示している。何かをインフラとして扱うということは、それを無色透明な、あたかも存在しないかのように扱うということの意味している。それに対してこれまで議論してきたブラックボックス化とは、何かを実現している技術や仕組みの存在は意識しても、その中身を問わない態度のことである。一見真逆の表現のようにみえるが、その本質は類似している。水道水の例でいえば、浄水システムの機構について（その存在を認知しながらも）関心をもたないことがブラックボックス化であり、その延長として、そもそも浄水システムのような仕組みの介在自体を意識しないことがインフラ化である、と整理できるだろう。

このとき私たちは、なぜ水道水を飲用したり料理に使用したりできるのかといえば、水道システムというインフラに対する信頼が確立しているからである。実際に取水施設の状況や安全対策の遂行について確認をしなくても、水道システムが安全に運営され裏切られることがないだろうという推測こそが、信頼の効用である。ニコラス・ルーマンは、このような信頼の機能を、社会的な複雑性を縮減し、いわば私たちの意志決定のコストを下げる態度として位置づけている。ルーマンの言葉を借りれば、信頼とは「過去から入手しうる情報を過剰利用して将来を規定するという、リスクを冒す」ことである⁵。ルーマンはまた「信頼の基礎には幻想がある。そもそも〔信頼する場合には〕上首尾に行為するために必要な情報が、与えられていないのである。行為する者は、欠けている情報をあえて無視する」とも述べている⁶。

このことはまさに、人々がブラックボックス化したシステムをインフラとして扱うことの本質を指し示しているといえるだろう。これまで論じてきた通り、ブラックボックス化とは、複雑な内的構造を（敢えて）無視するということであり、それはすなわち、内的構造の検証結果や動作の過程に対する確証がなくても、外的にそれが順調に作動しているという過去の事実を「過剰に」評価することで、それが将来にわたって問題なく動作すると信じることに他ならないからだ。つまり、一見逆説的ではあるが、何かを信頼してその将来に期待をするということはすなわち、その信頼の対象に対して不確実性が存在していることが前提なのである。山岸俊男は、この現象を「信頼のパラドックス」と呼び、「社会的な不確実性がまったく存在しない状況では、信頼は果たすべき役割を持たない」と論じている⁷。

CRDS（国立研究開発法人科学技術振興機構研究開発戦略センター）はこの「信頼（Trust）」の存立構造について、ルーマンや山岸らを含む人文・社会科学的なアプローチと、情報科学によるアプローチを幅広くレビューし、「信頼」という社会的現象を信頼する側（Trustor）と信頼される側（Trustee）の二者関係としてモデル化している⁸。特に重要な知見は、Trustor 側の「期待」である「信頼」とはあくまで Trustor の主観に基づくということである。なぜなら信頼とはこれまで論じてきたとおり、信頼の対象である Trustee に対して、確証がないにもかかわらず「裏切られることがないことを期待」することに他ならないからだ。そしてその対象となる Trustee 側がもつ信頼されうるための特性が「信頼性（Trustworthiness）」である（図 1）。

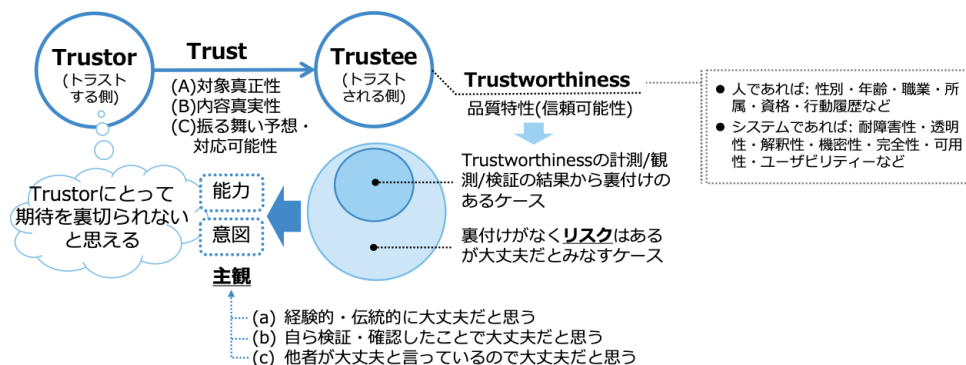


図 1. 信頼(Trust)のモデル図⁹

このモデルにおいて「信頼」とは、対象の「信頼性」についての観測・検証の結果から裏付けのあるケースを手がかりとしつつ、Trustee の能力や意図についての将来的な可能性を推測して、リスクをとっても大丈夫だとみなす範囲を主観的に判断することで成立する。この裏付けのあるケースと、裏付けのない（リスクをとる）ケースの比率がどの程度なら大丈夫だと思うのかは、Trustor の主観によって異なる。

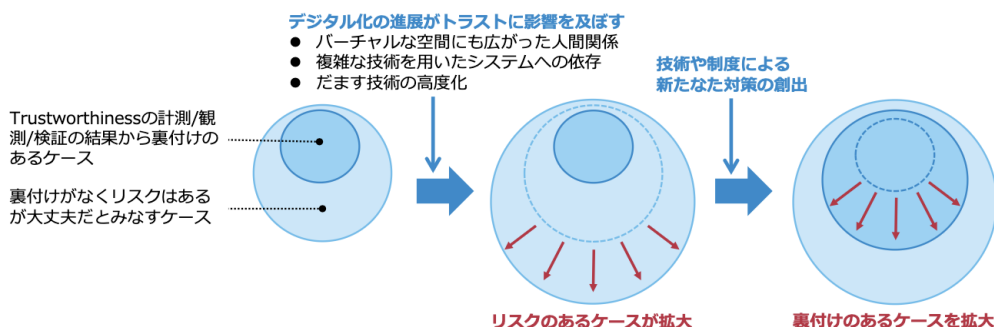


図 2. 信頼におけるリスク¹⁰

そして、特にソーシャルメディアなどのデジタルメディアにおいては、この「信頼性」の観測・検証可能性が十分担保されておらず、裏付けがなくリスクのあるケースが拡大する傾向があると指摘する（図 2）。これまで論じてきた通り、私たちはアテンション・エコノミー環境において、ソーシャルメディアの仕組みをブラックボックスとして扱い、いわばそのリスクを無意識のうちに過小評価して過剰に信頼している状態にあると考えるべきだろう。CRDS によれば、そのリスクを軽減し信頼の裏付けを強化するには、複数の情報源、複数のメディアによる「社会的よりどころ」の複合化が重要とされる

¹¹。

また、CRDS では「信頼の 3 側面」として、相手の何を信頼するのかについて 3 つの分類を提示している。ひとつ目は「対象真正性」であり、相手が本人・本物であるか、という問いに対応す

る。ふたつ目は「内容真実性」であり、やり取りする内容が事実・真実であるか、という側面である。そして3つ目は「振る舞い予想・対応可能性」で、対象の振る舞いが想定可能であるか（という意味で裏切りがないか）というものである¹²。プラットフォームのコミュニケーションにおいてこのうち「対象真正性」と「内容真実性」は、やり取りされるコンテンツの次元における信頼を問うものと位置付けられる。たとえば、プラットフォームに接続する他のユーザーが、なりすましなどをしていない本物であるか、が「対象真正性」であり、そのユーザーが投稿した内容に虚偽がないか、が「内容真実性」である。一方の「振る舞い予想・対応可能性」は、コンテンツを流通させるメディアの次元における信頼を問うものである。たとえば、プラットフォームが、検索キーワードに対して適切な結果を一貫して返答するかどうかや、タイムラインに表示されるコンテンツが期待した通りの順序で表示されるかどうか、などが相当する。

ソーシャルメディアの信頼を考えると、この二重の次元（コンテンツの次元とメディアの次元）を区別しておくことは極めて重要である。たとえば偽・誤情報の流通においては、当然ながらそのコンテンツ自体が「真実」であるかが疑われ、その発信者が「真正」であるかが問われるべき事態であるが、このコンテンツの次元の信頼を維持・回復するためには、メディアの次元における信頼が担保されていることが前提となる。しかしこれまで論じてきたとおり、ソーシャルメディアの「振る舞い」は、その信頼性が裏付けのあるものとして信頼されているというよりも、その裏付けが明確でないままにブラックボックスとして扱い、検証することを放棄することで、信頼できるかのようにやりすごしているという現実がある。

山岸俊男は、相手が裏切る可能性についての不確実性が存在している状態で、その相手の意図や行動に期待をすることを「信頼」とよび、相手が裏切ることがまずない（不確実性が小さい）と認識している状態で、その相手の意図や行動に期待することを「安心」とよんで区別した¹³。山岸のいう「安心」は、強い紐帯に基づくコミットメント関係を固定的に構築することによって、不確実性を低減することを指す。ソーシャルメディア上の（コンテンツ次元における）「信頼」は、本来は裏切られることに対する不確実性を「敢えて無視する」というメタ的な態度であるはずだが、現実にはその「信頼」を「安心」にすり替えたいとする社会的な欲望が強いのではないだろうか。すなわち、エコーチェンバー内において、意見の同質性が高いコミュニティと相互作用を継続することは、「裏切られる不確実性」を低減し、快適な「安心」に類似した環境を構築しうることになる（そしてこの「安心」と「依存」が表裏一体であることはいままでの議論）。しかしこの「安心」の論理は、「信頼」の裏付けとなる「信頼性」を対象の性質ではなく、対象との関係性（コミットメント）に求めるものであり、ひとたびエコーチェンバーの外部に出ればその裏付けが失われてしまうものだ。このように考えると、信頼の対象（Trustee）の信頼性そのものについての「社会的よりどころ」を回復し、さらに複合化するような、メディアの再設計が急務と言えるだろう。

コンテンツ次元における信頼の裏付け拡大に向けた取り組みは、進んでいないわけではない。たとえば、Xは「コミュニティノート」という機能によって、特定の投稿の信頼性に疑義がある場合、他のユーザーがそれを指摘しその議論を表示することで、集合知的にその信頼性を高める試みが始まっている。根拠のない書き込みに対して根拠のない反論が交わされるなど、その有効性については批判もあるが、とくにコンテンツ次元における「内容真実性」について、「社会的なよりどころ」を複合化するという点では、一定の意味があるといえるだろう。また、複数のファクトチェック団体による、ファクトチェック活動も、流通する情報量に対してチェック可能なリソースが少ないという問題を指摘されつつも、やはり「社会的なよりどころ」を増加させる取り組みとして重要だろう。これらのコンテンツにかかわる対象真正性、内容真実性をめぐっては、偽・誤情報の問題にアテンションが集まるにつれてその対策にも社会的な期待の圧力が高まっており、その信頼性の裏付けの少なさに対する問題意識は共有されつつある。

しかし一方で、メディア次元の「振る舞い予想・対応可能性」における信頼については、その信頼性の裏付けが希薄であることについての問題意識自体、必ずしも明確化していない。私たちは、アルゴリズムによるメディアの「振る舞い」について、何らかの予想や期待をもつことすらなく、「振る舞われるまま」に行動させられているともいえる。たとえば、ソーシャルメディアがスマートフォン上に

「通知」を表示するという「振る舞い」によって、私たちは日常生活の動作を止め、その「通知」に従って投稿を確認し、場合によってはそのままタイムラインのおすすめの「振る舞われるまま」にソーシャルメディアを使いつづけてしまう。その際にアテンションの対象となるのは、（メディアではなく）そこに表示されるコンテンツである。注意すべきなのは、コンテンツ次元における倫理の遂行（コミュニティノートなどの信頼性のよりどころ増加）をプラットフォームが担うように見せかけることによって、メディア次元の「振る舞い」への主観的な信頼はむしろ増す効果をもつということだ。しかしこのことは、「振る舞い」そのものの信頼性に対するよりどころが増加することを意味してはしない。

プラットフォームの「内部」からの信頼性確保の取り組みに期待することは当然ながら限界がある。まさに「社会的よりどころ」を確保するためには、「外部」からどのような規制や圧力がありうるのか、という議論がより重要といえるだろう。そこでは法的・制度的な規制もありえるが、利用者であるユーザー自身が、なにをもって信頼の裏付けとするのか、ソーシャルメディアにどのような「振る舞い」を期待し、それが裏切っていないことをどのように判断するのか、を問い直すことも重要だろう。なにが真実なのか、というコンテンツ次元の信頼だけでなく、メディアがどのように振る舞うべきなのか、というメディア次元の信頼をどのように確立するのか、メディアのデザインに異議申し立てするようなユーザーからの社会的な圧力こそが、ソーシャルメディアの倫理的なデザインを可能にする原動力となるのだ。アテンション・エコノミーを相対化し、情動的な反応ではなく、理性的な判断・推論に基づく倫理的なコミュニケーションを可能にするには、そのようなしくみの振る舞いに対する社会的な期待値のすり合わせが必要である。エコーチェンバーの「安心」を抜け出し、理性に基づく信頼が機能するようなメディアを再構築していくために、ブラックボックスを相対化し、その中身を問うような議論を深めていくことが重要だろう。

<執筆者プロフィール>

専門分野：メディア論、メディア・リテラシー

所属：武蔵大学社会学部准教授、武蔵学園データサイエンス研究所員

学位：博士（学際情報学）東京大学

主な著書：『アルゴリズム・AIを疑う 誰がブラックボックスをつくるのか』（集英社、2025年）、『Google SEOのメディア論 検索エンジン・アルゴリズムの変容を追う』（青弓社、2025年）など

-
- ¹ Latour, B., Pandora's Hope: Essays on the Reality of Science Studies, Harvard University Press., 1999, p.356
- ² Twitter, "Twitter's Recommendation Algorithm", 31 March 2023 (Retrieved on Jul 16, 2025; https://blog.x.com/engineering/en_us/topics/open-source/2023/twitter-recommendation-algorithm)
- ³ Google, 2025 (Retrieved on Jul 8, 2025; <https://www.google.com/search?q=X+アルゴリズム>)
- ⁴ Star, S. & Ruhleder, K., Steps Toward an Ecology of Infrastructure: Design and Access for Large Information Spaces, Information Systems Research, 7(1), 1996, pp.111-134.: 113
- ⁵ Luhmann, N., Vertrauen : ein Mechanismus der Reduktion sozialer Komplexität. UTB, 1973 (大庭健・正村俊之訳『信頼：社会的な複雑性の縮減メカニズム』勁草書房、1990 年、33 ページ)
- ⁶ Luhmann, N., op. cit. (大庭健・正村俊之訳 前掲書、55 ページ)
- ⁷ 山岸俊男『信頼の構造: 心と社会の進化ゲーム』東京大学出版会、1998 年、15 頁
- ⁸ CRDS（国立研究開発法人科学技術振興機構 研究開発戦略センター）「戦略プロポーザル デジタル社会における新たなトラスト形成」、2022 年、8-9 頁
- ⁹ CRDS、前掲論文、9 頁
- ¹⁰ CRDS、前掲論文、11 頁
- ¹¹ CRDS、前掲論文、11-12 頁
- ¹² CRDS、前掲論文、1-2 頁
- ¹³ 山岸俊男 前掲書