

社会を含む広義の人工物に対する知識の構築と運用論

出口弘 (Hiroshi Deguchi)
東京工業大学

1 人を含むシステムに介入する知識の構築と運用の課題

人や社会を対象とした知識の構築と運用法は、物理的な対象に対する知識の構築と運用法と大きく異なることは、自然科学を対象とする科学論ではほとんど言及されていない。社会科学の領域でその知の構築と運用を論じる際にも、自然科学的な知の構築と運用の方法を規範的な方法として援用することが多い。しかし人や社会を対象としてそこに何らかの意図を持って介入するための知識の構築と運用が、自然科学とは異なる方法論を必要とすることもまた社会科学では古くから認識され議論されている。社会という人と組織・制度と「もの」と情報システムからなる複合的人工物には、ビルや橋梁のような「もの」として物理的な法則に従う人工物（以下物理的人工物）が含まれる。「もの」としての人工物を設計するには、物理的な知識あるいはそれを特別な境界条件で制約した性質に関する知識が必要とされる。自然物を加工した物理的人工物は通常、橋梁やビルの設計のように、何らかの利用目的（外部目的）があり、それに基づき物理学的な知識を用いて構築をする。何らかの外部目的に基づき製作された物理的人工物はそれが構築されることで社会に何らかの形で介入し影響を与える。それは関連する社会のステークホルダー自身の目的（内部目的）とコンフリクトを生じることがあり、そこではステークホルダー間の合意形成という人工物の運用上の課題が生じる。この「もの」としての人工物に、さらに情報システムが加わり、制御や最適化の概念が入ることで新たなコンフリクトが生じる。制御を含む人工物（人工システム）では、システム内部に設計者の与える目的（外部目的）を制御目標として組み込む必要がある。これはビルや橋梁のような人工物にはない性質である。制御や最適化により介入する対象が、人を含むシステムであるとき、制御対象となる主体自体の持つ目的（内部目的）と制御を行う人工システムの中で設定された制御目標との間にコンフリクトが生じる可能性がある。人工システムに組み込まれた制御目標が、対象となる人の内部目的やそれに基づく行為を変化させることを目標とする場合、その介入は合意形成の視点からも倫理上も大きな問題をはらむ。この社会や個人に対する制御を行う人工システムでは、介入側の制御目標を介入される側に対して開示して合意することなく、あるいはそもそも制御目的そのものをそれにより介入される主体に隠した形で設計され介入することがしばしばである。このような「他者の意図や目的そのものを制御しようとする介入」は近年急速に普及しつつある情報システムや人工知能の中で広く用いられつつある。これは現実に対して介入する人工物に対する知の構築と運用にさらなる方法論的課題を惹起する。

2 広義の人工物に対する知識の構築と運用に対する二つの誤謬

現実介入することを目的とする人工物に対する知を扱う場合、「科学」という言葉でラベリングされた知の構築と運用の多義性については細心の注意を払う必要がある。「Aという薬に効果があるか否かを問う」薬による人への治療介入の効果の判定に関する知の運用では、介入対象である患者の体内で薬が作用する因果的機序についてのモデルの構築とその真偽の判定は必ずしも求められない。現実への介入が、偶然ではなく効果があり、その介入が目的の実現に有効であることを示すための知の運用そのものには、因果的機序のモデルは不要である。そこでは介入の結果が目的からみて有効であるかの有効性(Efficiency)評価がまず求められる。その上で目的に基づいた対象への介入の過程で、目的や代替案がステークホルダーと合意形成が取れておりかつそこでの介入プロセスが妥当であるかという、妥当性(Adequacy)評価がなされることが求められる。その上で、これらの評価が介入プロセスの諸段階にフィードバックされることが求められる。ここでは、対象に対する因果的な機序に関するモデルという形での知の構築とその検証あるいは反証のための知の運用を「真偽のコード(でのコミュニケーション)に基づいた知の構築と運用」と呼ぶ。これに対して目的に基づいて現実介入するための知の構築と運用をここでは、「(合意形成と介入プロセスの)妥当性と(結果の)有効性のコード(でのコミュニケーション)に基づいた知の構築と運用」と呼ぶことにする。

目的に基づいて現実介入を行うための知の構築とその運用は、工学領域から社会科学まで、広範な領域で行われている。そこでは「妥当性と有効性のコードに基づいた知

の構築と運用」が行われねばならない。しかし多くの場合、それがきちんと認識されないまま、着目した代替案による介入が有効であることの統計的な判定(検定)に「科学」というラベリングがなされるという現状がある。これを人工物に対する知の構築と運用の**第一の誤謬**と呼ぶ。さらに近年大きく課題となりつつあるのが、制御する或いは最適化する人工システムによる、社会や個人に対する介入という問題である。制御の基本概念であるフィードバックは、「目的」という概念を陽に設定することなしには定義できない。したがってフィードバックの数理モデル自体に、設計者の設定した何らかの制御目標(外部目的)とそれに基づく介入概念が組み込まれている。これは最適化の数理モデルも同様である。このような制御・最適化の枠組みを、人や社会を対象して用いる際には、設計者により想定される外部目的は、制御目標としてシステムに組み込まれ、制御される対象となる人や社会に通常開示されない。この人工システムに組み込まれた「制御目標」や「制御された合意形成」について介入される側に明示的に示すことが主体を含む対象に対する介入では求められる。この制御目標の開示はビルや橋梁の設計の目的を開示することとは本質的に異なる。ここで課題となっているのはシステムに組み込まれた制御目標である。エアコンのような制御システムでは、システムに組み込まれた制御目標、例えばエアコンの温度設定は、エアコンの使い手にも開かれている。またエアコンの利用者の目的もまた温度管理であり、外部目的と内部目的は基本的に矛盾していない。というよりも温度を制御したいというエアコンの利用者の内部目的をシステムの制御目標として組み込んだ人工システムがエアコンである。従来の物理的人工物では、使い手の内部目的を遂行する作業は、使い手自身が行い、そのための道具として人工物を用いる。その人工物の仕様は使い手の内部目的を参照して設計者が決める。例えば「ボールペン」は書くためのものであり、「自動車」は移動するためのものであるという想定される「外部目的」とそれに基づいたボールペンの使用温度範囲、自動車の燃費や最高速度などの様々な仕様は設計者の側で定める。しかしボールペンを用いて筆記をする、あるいは車を運転するのはその人工物の使い手であり、そこに文字を書くあるいは運転するという内部目的があり、その手段として道具としての人工物を用いる。設計者はその意図を仕様化するという相互関係がある。だが制御や最適化を含む形での介入では、制御や最適化の対象が、主体を含むシステムとなる。このとき対象となるシステムに含まれる主体の目的を支援するというよりは、主体の目的遂行を取り上げそれをシステムが代替するあるいは、外部目的によって対象に含まれる主体の内部目的そのものを変えるという意味での介入を企図したシステムを設計と実装がなされる。このような介入側の外部目的と介入される側の内部目的にコンフリクトが生じ得ることを、人工物に対する知の構築と運用の**第二の誤謬**と呼ぶ。

3 結語

人を含む対象へと「介入」する知の構築と運用では、介入対象相手の内部的な意図・目的・代替案選択(意思決定)・合意形成への、合意なき干渉を問題とせざるを得ない。あらゆる社会的活動は他者への介入を含む。それゆえ介入を含む知の構築と運用では、システム実践やアクションリサーチから公衆衛生・医療領域さらに社会工学まで、様々な形で介入対象の主体と介入する側の合意形成プロセスや倫理基準を問題としてきた。これは橋梁やダムのような物理的人工物の構築でさえ例外ではない。橋梁やダムはそれ自体は物理的な構築物でありそこでの介入目的は介入される側の目的と全面的矛盾するものではないが、環境アセスメントを含む関係者の合意形成のもとでの社会的実装プロセスが求められている。医療では治療により介入される患者と介入する医者の間でのインフォームドコンセントが要請される。さらに医療実験ではヘルシンキ宣言に基づく倫理基準が定められている。これに対し制御や最適化を含む技術が主体あるいは主体の活動へ介入する際には、対象とのインフォームドコンセント以上の問題が生じる。これは個人のレーティングによる行動の矯正や個人向けにカスタマイズされた情報による政治的意思決定への干渉という露骨な介入だけの問題ではない。例えば自動運転では、センサーで認知したリスクに対してどのような意思決定をするかに関し、制御目標や意思決定原理の選択がシステムに組み込まれており、運転者はそれに関与できない。制御を含むシステムは制御目標を組み込む人工物であり、設計者が設定しシステムに組み込んだ制御目標や制御の方法が、介入され制御される対象の主体にとって隠されている、あるいは介入する側とされる側の目標が対立することがあることを認識する必要がある。介入される主体の内部目的を制御・最適化するメカニズムが人工システムに組み込まれる得ることと、そのリスクを人工物の科学論として認識し論じる事が求められる。