

椅子や机を運ぶとき、何かぶつけて思わず「いてっ」と言ってしまったこと、ありませんか。人間はどうも他者の痛みを自分のものとしがちな動物のようで、逆に、困っている人に共感せず助けないと「人でなし」なんて言われかねない。でも困っている人がいたら必ず助けるべきかっていうと、これはけっこう難しい。例えば悪い人が困っている時に助けるのは、果たして良いことなのか。悪い人が困っている時に助けなかった人が困っている時に助けることは？ そもそも良いとか悪いとかって何なの？

この難問を「情けは人のためならず」というフレームから議論してきた人たちがいます。他人を助けると巡り巡っていずれは自分に返ってくると諺は言うけれど、そんなこと本当に実現しうるのか。前提を絞ると数学的に研究できると言うのです。助けることのコストはどれくらいで、助けてもらうことの利益はいかほどか。善悪の基準はどうなってるか。諸々を設定して計算する。そう、Ohtsuki & Iwasa (2004; 2006) みたいに。すると「善人は善人を助ける」「善人を助けない人は悪人」「悪人は助けない」「悪人を助けなくても悪人とされない」「善人を助ける悪人は善人」という条件下で「情け」が安定して巡ると分かる。何がすごいって、この納得感ある人間臭い言説が数式の解析だけから導出されているのです。

そんなわけで個人的お気に入りベスト20に入る研究なわけですが、そう上手く行く場合だけじゃないよ、という論文が出たようなので読んでみました (Hilbe et al., 2018)。先ほどの計算には、ある人 (Xさん) への評価は社会で共有されているとの前提がありました。これを、人によってXさんへの評価は異なるものとする。そうすると話が複雑すぎて解析的には解けないのでシミュレーションしたそうです。コンピュータの中に何人もの仮想ニンゲン (エージェントと呼びます) を作る。各エージェントは、いつ助けるか、何を善または悪とするか、

各々なりのルールを持っています。それらが相互作用する仮想社会を作って、どういうエージェントの集まりなら「助け合いの輪」が循環するか、試してみた。

どうなったか。Xを悪人と思っているAさんが居たとしましょう。AはXを助けません。するとXを善人と思ってるBはAを悪人認定します。ところがAの行為を見てなかったCは、引き続きAを善人と思っています。そのためAを助けないBは、Cからすると悪人になってしまふ。それでCがBを助けないと、DがCを悪人認定し……という地獄のような状況が発生しかねないというのです。

衝撃的だったのが次の段落。そういうケースをより詳しく調べるために、全員が全員を善人と思っている中で、一人だけが、間違っている一人を悪人認定した場合に何が起きるか検討したというのです。自分がその「勘違いエージェント」になった場合を想像して絶望的な気持ちになりました。仮想社会の“神”から「あやまち」を負わされ、ために皆から爪弾きにされ、果ては「助け合いの輪」崩壊の元凶となってしまうエージェントが気の毒すぎる……。

気持ちを強くして「詳しくはこっちを」と書かれた補足資料を読みました。真理の追求に捧げられたエージェントの行末を見届けねばならないと思ったのです。そしたら拍子抜け。なんのことはない、一人だけが誤解している場面ならシミュレーションを回す必要はなくて、数式の解析でOKだったんですね。犠牲にされた仮想エージェントは存在しなかったのです。心底安堵しました。いかに科学のためであっても、犠牲は最小限に留めるべきですよ。



## Profile — 平石 界

東京大学大学院総合文化研究科博士課程退学。東京大学、京都大学、安田女子大学を経て、2015年4月より現職。博士 (学術)。専門は進化心理学。