

SYLWETKA ERNSTA FEHRA

Michał Krawczyk
Uniwersytet Warszawski

Ernst Fehr to badacz pod wieloma względami nietuzinkowy. Jest jednym z najbardziej wpływowych przedstawicieli ekonomii eksperymentalnej i behawioralnej. Jego pionierskie badania w obszarze neuroekonomii każą uznawać go za współtwórcę tej subdyscypliny. Wreszcie, jako jedna z niewielu światowych supergwiazd profesji, pracuje poza Stanami Zjednoczonymi.

ŻYCIORYS

Ernst Fehr urodził się w miasteczku Hard nad Jeziorem Bodeńskim w roku 1956. Będąc człowiekiem wielu talentów i niewątpliwej ambicji, porzucił doskonale rozwijającą się karierę zapaśniczą (sic), by skupić się na ekonomii. Doktorat i habilitację uzyskał na Uniwersytecie Wiedeńskim, a od roku 1994 jest profesorem Uniwersytetu w Zurichu; przyznano mu także tytuł *Global Distinguished Professor* na Uniwersytecie Nowojorskim. Publikuje nie tylko w czołowych pismach ekonomicznych, takich jak *AER*, *QJE*, *Econometrica* czy *Journal of Political Economy*, ale także w *Science*, *Nature* i *Neuron*. Jest laureatem licznych prestiżowych nagród, członkiem kilku akademii nauk, byłym prezydentem *Economic Science Association* i *European Economic Association*. Już te osiągnięcia, w połączeniu z powracającymi jak bumerang pogłoskami o rychłej nagrodzie Nobla, wystarczyłyby zapewne, by większość środowiska zółkła z zazdrości. Czarę goryczy przelał jednak stumilionowy grant banku UBS. Te uczucia tłumaczą zapewne dość niekonkretne i słabo udokumentowane zarzuty nieuczciwości akademickiej, które ciekawi łatwo znajdą wśród złośliwych plotek na temat innych ekonomistów na portalu www.econjobrumors.com. Spektakularny sukces Fehra stoi też być może za krytyką niektórych jego osiągnięć naukowych, w szczególności, o ironio, modelu uwzględniającego uczucie zazdrości, o czym za chwilę.

Pozwolę sobie z każdego z głównych obszarów twórczości naukowej Fehra wybrać co najwyżej kilka najszerzej dyskutowanych artykułów. Kompletny i szczegółowy przegląd całej jego dotychczasowej spuścizny wymagałby bowiem o wiele dłuższego wywodu, niż redakcja byłaby gotowa przepuścić (a czytelnik – strawić). Wynika to między innymi z tego, że – przeciwnie do wielu badaczy piszących całe życie nowe

wersje tego samego artykułu czy książki – Ernst Fehr przejawia niezwykłą otwartość na nowe koncepcje i techniki badawcze, więc też i jego dorobek jest zróżnicowany. Nie oznacza to wszakże, że nie daje się znaleźć w nim wątków wiodących. Dostrzegam tutaj dwa następujące.

Po pierwsze, znacząca część jego artykułów zainspirowana jest zagadkową skłonnością homo sapiens do współpracy, kosztem własnego (krótkoterminowego) zysku. Istotnie, wbrew temu, co można by sądzić, patrząc na telewizyjne wiadomości ukazujące na ogół konflikty zbrojne, utarczki polityków czy choćby szalone wyścigi konsumentów walczących o przeceniony sprzęt elektroniczny (a ostatnio: nawet o papier toaletowy), ludzie są w stanie, ogólnie rzecz biorąc, organizować się w sposób wyjątkowo harmonijny. Oczywiście o tym, dlaczego tak jest, napisano już całe biblioteki, ale dużo odpowiedzi ma zwykle ten, kto nie ma jednej dobrej. Chęć rozprzestrzeniania własnych genów może tłumaczyć współpracę, ale tylko z krewnymi, wzajemność może ją tłumaczyć, ale tylko w powtarzalnych interakcjach, budowanie reputacji - tylko gdy interakcje są nieanonimowe itd. Tymczasem u homo sapiens obserwujemy rutynową współpracę w obrębie ogromnych, niespokrewnionych grup, w jednorazowych, anonimowych interakcjach.

Drugą osią spinającą twórczość Fehra jest konsekwentne wykorzystywanie badań eksperymentalnych. Choć mogłoby się wydawać, że proste, abstrakcyjne gry laboratoryjne nie mają wiele wspólnego z ewolucją ludzkiej zdolności do współpracy, to jednak właśnie one pozwalają odróżnić poszczególne mechanizmy, które współpracę tę umożliwiają (lub utrudniają).

AWERSJA DO NIERÓWNOŚCI

Jeden z takich mechanizmów analizuje najczęściej cytowany artykuł Fehra (Fehr i Schmidt, 1999). Badacze proponują model awersji do nierówności, zakładając, że niektórzy decydenci nie są samolubni, ale odczuwają dyskomfort, gdy ktoś z ich grupy odniesienia jest bogatszy lub biedniejszy od nich. Te preferencje można kojarzyć, odpowiednio, z zazdrością i poczuciem winy, przy czym nie muszą one być zresztą równie intensywne: większość osób sytuację, w której ktoś inny ma więcej, odczuwa jako bardziej niesprawiedliwą. Naturalnie, koncepcja taka nie jest w naukach społecznych nowa, traktują o niej choćby prace Festingera, ale w ekonomii była w dużej mierze ignorowana – homo oeconomicus jest bowiem istotą samolubną.

Osiągnięciem Fehra i Schmidta było sformułowanie prostego modelu awersji do nierówności, który łatwo można zastosować do eksperymentów behawioralnych. Co istotne, w wielu z nich – takich jak gra przetargu ułtymatywnego, gra w dobra

publiczne czy gra w zaufanie/wymianę dóbr – model Fehra-Schmidta daje w miarę poprawne predykcje. Przykładowo, w grze w zaufanie pierwszy gracz decyduje, jaką część początkowego zasobu przekazać drugiemu (który zaczyna z niczym). Kwota ta zostaje pomnożona przez eksperymentatora (np. przez trzy), po czym drugi gracz decyduje, jaką jej część odesłać z powrotem pierwszemu. Samolubny drugi gracz zachowa wszystko dla siebie, więc stosując indukcję wsteczną stwierdzamy, że w jedynej równowadze stabilnej względem podgier pierwszy gracz nie przekaze drugiemu ani złotówki. Jeśli jednak pierwszy gracz sądzi, że drugi ma wystarczającą awersję do posiadania więcej niż inni, to będzie oczekiwać uczciwego podziału pomnożonej kwoty. W jego własnym interesie jest więc wysłać drugiemu pieniądze, tym bardziej jeśli sam odczuwa dyskomfort z powodu posiadania więcej od innych (który to przypadek zajdzie, gdy zachowa cały zasób początkowy dla siebie).

W praktyce większość graczy (obsadzonych w każdej z ról) jest skłonna się dzielić. Jest to zachowanie zgodne z koncepcją awersji do nierówności, a sprzeczne z powszechnym egoizmem. Warto podkreślić, że model wcale nie przewiduje współpracy w każdych okolicznościach. Przykładowo: dopuszczenie „rynkowej” konkurencji między proponującymi w grze przetargu ultymatywnego prowadzi, mimo awersji do nierówności, do (trafnych) przewidywań zbieżnych z przewidywaniami modelu z samolubnymi graczami.

W swej bazowej postaci model Fehra-Schmidta generuje natomiast mniej celne przewidywania dla gry w dyktatora: wynika z niego bowiem, że dyktator powinien albo podzielić „ciastko” po równo, albo zatrzymać je w całości dla siebie. W rzeczywistości znacząca część badanych wybiera którąś z opcji pośrednich; otrzymanie tego wyniku wymagałoby modyfikacji założenia o liniowości (podobnie jak to uczyniono w konkurencyjnym modelu Boltona i Ockenfelsa, 2000). Mamy tu do czynienia z typowym wyborem: prostszy model daje bardziej jednoznaczne przewidywania, ale oznacza to, że łatwiej go sfalsyfikować.

Większe znaczenie mają dwa inne nurty krytyki modelu Fehra-Schmidta. Po pierwsze, zarzucano im nazbyt swobodne cytowanie własnych przewidywań, tak by brzmiały mocniej i lepiej pasowały do nowych danych. Najbardziej wymownym krytykiem był tu Shaked (2005): *“When I finally came to read [Fehr and Schmidt’s inequity aversion papers], I was introduced to a subculture hitherto unknown to me, a subculture that apparently coexists parallel to main-line economics and in which different rules of logic and different laws of proof apply. In this sphere it seems to be permitted to misquote one’s own theorems, to place crucial information in appendices and footnotes, to treat data in a casual and nonchalant way, and it is allowed to inflate results when citing them in subsequent papers.”*

Po drugie, badania eksperymentalne pokazały, że teoria zakładająca – jak model Fehra-Schmidta – że użyteczność zależy wyłącznie od końcowych wypłat, które osiągną uczestnicy interakcji, w wielu grach daje błędne przewidywania. Wiele osób może akceptować nawet znaczne nierówności, gdy są one dziełem przypadku, ale nie skutkiem (samolubnych) wyborów dokonanych przez partnera interakcji (zob. np. Blount, 1995). Te wyniki zainspirowały dwa kierunki badań. W jednym z nich skupiono się na równości *szans* otrzymania wysokiej wypłaty (a zatem równości *oczekiwanych* wyników). Stosownie do tego sformułowano tzw. „proceduralne” wersje modelu Fehra-Schmidta (Trautmann, 2009, Krawczyk 2011).

Drugi rodzaj reakcji na ograniczenia modelowania użyteczności zależnej tylko od końcowych wyników to dopuszczenie, że w wielu sytuacjach zasadnicze znaczenie mają postrzegane *intencje* partnera interakcji. Są to modele wzajemności (*reciprocity*) jak choćby ten Falka i Fischbachera (2006). Dobrze o Fehrze świadczy, że zamiast trzymać się swojego popularnego modelu jak umarły kadzidła, potrafił wnieść istotny wkład także do nurtu literatury wykorzystującego to alternatywne podejście.

WSPÓŁPRACA I SANKCJE

Głównym wehikulem wykorzystywanym tu przez Fehra i współpracowników (głównie Simona Gächtera i Ursa Fischbachera) była gra w dobra publiczne. W jej klasycznej formie, w każdej z np. 10 rund, każdy z n (najczęściej czworga) członków grupy decyduje, jaką część swojego początkowego zasobu (np. 10 złotych) zatrzymać dla siebie, a jaką przekazać na wspólny rachunek. Ten zostaje przez eksperymentatora pomnożony i rozdzielony po równo pomiędzy członków grupy. Wartość mnożnika jest przy tym większa od 1, ale mniejsza od n , tak że łączny dochód grupy jest maksymalny, gdy wszyscy przekażą całość początkowego zasobu na wspólny rachunek, ale chcąc maksymalizować swoją własną wypłatę, każdy z graczy winien zachować cały zasób dla siebie, niezależnie od tego, jakiego zachowania pozostałych się spodziewa. Mamy więc do czynienia z klasycznym dylematem współpracy – byłoby wspaniale, gdyby wszyscy płacili uczciwie podatki, sprząтали po swoich psach itd., ale – wobec niewielkiego prawdopodobieństwa sankcji – wiele osób nie ma na to ochoty.

Eksperymenty oparte na grze w dobra publiczne pokazują, że choć początkowo zwykle średnie kontrybucje do wspólnego rachunku są znaczące, w kolejnych rundach najczęściej spadają. Jedno z wyjaśnień jest takie, że ludzie są skłonni do współpracy, ale tylko warunkowo – tj. chcą współpracować tylko jeśli inni współpracują (nb. Fischbacher et al., 2001, potwierdzili to w eleganckim badaniu opartym na „me-

todzie strategii” Seltena, 1965). Jeśli więc w grupie trafi się choć jedna „czarna owca” – pozostali równają w dół.

Fehr i współpracownicy przeciwstawili tej klasycznej wersji gry w dobra publiczne grę z możliwością kosztownego, indywidualnego karania „jadących na gapę” członków grupy. W każdej rundzie, każdy z uczestników, dowiedziawszy się, jaką część początkowego zasobu przekazał na wspólny rachunek każdy z pozostałych, może każdemu z nich przydzielić wybraną przez siebie liczbę „punktów karnych”. Każdy przydzielony punkt karny kosztuje karzącego, ale jeszcze drożej (np. trzykrotnie) – karanego. „Życiowy” ekwiwalent takich prywatnych sankcji to np. zwrócenie uwagi komuś, kto śmieci na ulicy – wiąże się to z pewnym ryzykiem wdania się w pyskówkę itp., ale jeśli faktycznie zniechęci to do śmiecenia, to z punktu widzenia społeczeństwa może być to korzystne (stąd możemy mówić o *altruistic punishment*).

Okazuje się, że znacząca część ludzi jest skłonna, mimo powstających kosztów własnych, karać gapowiczów, tj. przydzielać punkty karne tym, którzy nie przekazali wiele na wspólny rachunek. Naturalnie, bardziej skłonni karać w ten sposób są ci, którzy sami zachowali się hojniej. Szczególnie interesujące jest to, że znacząca gotowość do karania samolubów utrzymuje się nawet w grach z tzw. *stranger matching*, tj. gdy skład grupy zmienia się co rundę, zatem nie można liczyć na prywatny pożytek z „nawrócenia” ukaranych gapowiczów w przyszłości.

Oczywiście, gdyby poziom wpłat na wspólny rachunek nie zmienił się, możliwość karania obniżyłaby jedynie dodatkowo łączną wypłatę. Jest jednak inaczej: przeciwieństwo do klasycznej gry w dobra publiczne, współpraca wzmacnia się tylko w kolejnych rundach. Dlaczego ten mechanizm, choć sam w sobie też mający charakter dylematu współpracy, okazuje się efektywny? Sądzę, że tego rodzaju pytania skłoniły Fehra do badań w obszarze neuroekonomii.

NEUROEKONOMIA

Większość ekonomistów zdaje sobie sprawę z istnienia mózgu, ale traktuje go jako „black box” – nie zajmuje ich pytanie, w jaki sposób przetwarzana jest informacja i podejmowana decyzja. Fehr należy do pionierów podejścia alternatywnego. Przykładowo w artykule Knoch i innych (2006) pokazuje, że czasowe zaburzenie – przy pomocy przezczaszkowej stymulacji magnetycznej – funkcjonowania grzbietowo-bocznej kory przedczołowej (*dorsolateral prefrontal cortex, DLPFC*) w prawej półkuli redukuje skłonność do odrzucania niesprawiedliwie niskich ofert w grze przetargu ultimatywnego. W oparciu o wcześniejszą wiedzę na temat roli, jaką odgrywa ten obszar mózgu w podejmowaniu decyzji, można przypuszczać, że jest to związane z tym,

że stymulacja ta utrudnia odrzucenie pokusy wybrania opcji korzystniejszej dla decydenta. W innym głośnym artykule (De Quervain et al., 2004) Fehr i współautorzy badają podstawy altruistycznego karania w kontekście gry w zaufanie. Stwierdzają, że realizacja kosztownej dla karanego (nie symbolicznej) sankcji pobudza prądkowie grzbietowe, co może świadczyć o odczuwanej przez karzącego satysfakcji.

Ekonomista głównego nurtu może na to oczywiście wzruszyć ramionami. Badany nakłada na postępującego niesprawiedliwie kolegę sankcję, bo daje mu to satysfakcję – to niemalże tautologia. Nie zbliża nas to szczególnie do rozwikłania zagadki, dlaczego homo oeconomicus wyewoluował w tak łagodnej wersji, nieskorej do napadania na niespokrewnionych przedstawicieli swego gatunku bez dania racji, jak to chętnie czynią nasi najbliżsi krewniacy. Przypuszczam, że to chęć bardziej bezpośredniego podjęcia zagadki ewolucji kooperacji skłoniła Fehra do kolejnej wolty (choć innym powodem mogła być konstatacja, że tradycyjne badania neuroekonomiczne, polegające na przeskanowaniu mózgów trzydziestu badanych na krzyż straciły nieco swój blask, nie dostarczywszy ekonomii pożytków, na które liczone).

GENOEKONOMIA

Mimo zbliżającego się wieku emerytalnego Fehr nie zwolnił zatem, a rozpoczął przygodę z „genoekonomią” – zabawą jeszcze droższą od neuroekonomii, pozwalającą mu zrealizować przewagę komparatywną w postaci milionów otrzymanych od UBS. Wedle mojej wiedzy nie przyczyniła się ona jeszcze znacząco do rozwikłania zagadki współpracy, natomiast badacze potrafią np. zlokalizować geny, które mogą być współodpowiedzialne za gotowość do podejmowania ryzyka. Poprzednie badania pokazały, że skłonność ta jest w pewnej części dziedziczna (szczególnie cennego materiału dostarcza tu porównanie identycznych genetycznie bliźniąt z bliźniętami nieidentycznymi, Cesarini et al., 2009; można bowiem zakładać, że wpływ środowiska, ale nie genów, jest w każdym z tych przypadków jednakowy). Nie było jednak wiadomo, które konkretnie geny mogą być za tę dziedziczność odpowiedzialne.

W badaniu niemal miliona genomów, podpisanym przez niewiele mniejszą liczbę współautorów, w tym Fehra (Linner et al. 2019), zidentyfikowano 99 *loci* (obszarów chromosomu zajmowanych przez dany gen) powiązanych z deklarowaną ogólną skłonnością do podejmowania ryzyka. Z punktu widzenia naukowców społecznych (dla których fragmenty artykułu dotyczące regresji nierównowagi sprzężeń tudzież glutaminianu, kwasu g-aminomasłowego i innych neuroprzekazników mogą być mniej interesujące) najważniejsza jest chyba obserwacja, że znacząca część *loci* powiązanych z ogólną skłonnością do ryzyka powiązana jest także z jej korelatami, taki-

mi jak skłonność do nadmiernie szybkiej jazdy samochodem czy nadużywania alkoholu. Odkrycie takie sugeruje, że przynajmniej część wpływu zamiłowania do ryzyka na tego typu zachowania zdeterminowana jest genetycznie (a w rezultacie zapewne niemożliwa do wykorzenia przez edukację czy temu podobne interwencje). Jeśli Fehr i koledzy zdołają odkryć genetyczne podstawy kooperacji, to być może doczeka się on w końcu upragnionego telefonu ze Sztokholmu.

BIBLIOGRAFIA

- Bolton, G.E., & Ockenfels, A. (2000). ERC: A theory of equity, reciprocity, and competition. *American economic review*, 90(1), 166–193.
- Blount, S. (1995). When social outcomes aren't fair: The effect of causal attributions on preferences. *Organizational behavior and human decision processes*, 63(2), 131–144.
- Cesarini, D., Dawes, C.T., Johannesson, M., Lichtenstein, P., & Wallace, B. (2009). Genetic variation in preferences for giving and risk taking. *The Quarterly Journal of Economics*, 124(2), 809–842.
- De Quervain, D.J., Fischbacher, U., Treyer, V., & Schellhammer, M. (2004). The neural basis of altruistic punishment. *Science*, 305(5688), 1254.
- Falk, A. & Fischbacher, U. (2006) A theory of reciprocity. *Games and Economic Behavior*, 54(2), 293–315.
- Fischbacher, U., Gächter, S., & Fehr, E. (2001). Are people conditionally cooperative? Evidence from a public goods experiment. *Economics letters*, 71(3), 397–404.
- Knoch, D., Pascual-Leone, A., Meyer, K., Treyer, V., & Fehr, E. (2006). Diminishing reciprocal fairness by disrupting the right prefrontal cortex. *Science*, 314(5800), 829–832.
- Krawczyk, M.W. (2011). A model of procedural and distributive fairness. *Theory and decision*, 70(1), 111–128.
- Linnér, R.K., Biroli, P., Kong, E., Meddens, S.F.W., Wedow, R., Fontana, M.A., ... & Nivard, M.G. (2019). Genome-wide association analyses of risk tolerance and risky behaviors in over 1 million individuals identify hundreds of loci and shared genetic influences. *Nature genetics*, 51(2), 245–257.
- Selten, R. (1965). *Die Strategiemethode zur Erforschung des eingeschränkt rationalen Verhaltens im Rahmen eines Oligopolexperimentes*. Seminar für Mathemat. Wirtschaftsforschung u. Ökonometrie.
- Shaked, A. (2005). *The rhetoric of inequity aversion*. Available at SSRN 675227.
- Trautmann, S.T. (2009). A tractable model of process fairness under risk. *Journal of Economic Psychology*, 30(5), 803–813.