

POATE FI SALVAT *Utilitarismul* LUI MILL DE TEORIA JOCURILOR? *

J.R. LUCAS

Universitatea Oxford

The author offers a framework of analysis for Mill's *Utilitarianism* based on game theory, able to capture all the nuances of Mill's thought: that it is irrational to ignore the collective point of view, that it is better to keep to the rules than to try to perform the act that will have the best consequences etc. He believes that Mill would have welcomed this game-theoretical approach, although the problem of justice needs further arguments. Contrary to the static, solipsistic, exclusively future-oriented standpoint of the classical economists, we are forced, by re-thinking the case of utilitarianism, to recognise that rationality is dynamic, leading us to take a longer temporal and wider personal view of what is involved in the decisions we are called on to take.

Keywords: utilitarianism, game theory, Mill's proof, rationality.

Utilitarismul lui John Stuart Mill ne trezește interesul și ne provoacă simpatie pentru că e defect. El reflectă criza petrecută în viața lui Mill odată cu pierderea crezului pe care îl avea. Mill a fost crescut de către tatăl său în chingile celui mai sever utilitarism însă, din cauza lipsei de afecțiune, a suferit o depresie la începutul maturității. Utilitarismul ar fi putut să funcționeze ca ghid pentru buna guvernare a Indiei de către James Mill și colegii săi, dar era o hrană spirituală săracă pentru avântata mișcare romantică. El tratează oamenii ca unități, nu ca indivizi. Așa cum arată Bernard Williams în tranșanta sa critică a utilitarismului¹, acesta nu ia deloc în considerare "proiectele" pe care le urmăresc oamenii.

*Acest text a fost prezentat în cadrul Conferinței Internaționale "John Stuart Mill. 1806-2006", organizată la Universitatea din București, Facultatea de Filosofie, în perioada 3-4 noiembrie 2006. Pentru versiunea originală, vezi volumul *Proceedings of the International Conference "John Stuart Mill. 1806-2006"*, University of Bucharest, November 3-4, 2006; Ed. Valentin Mureșan & Cristian Ducu, Edit. Universității din București, București, 2007, pp. 9-19.

¹ J.J.C. Smart, Bernard Williams, *Utilitarianism: for and against*, Cambridge University Press, Cambridge, 1973.

Fiecare individ nu numai că are experiența durerii și plăcerii ca efect a ceea ce i se întâmplă, dar este de asemenea inițiatorul unor acțiuni conforme cu politicile pe care le-a creat în lumina viziunii sale despre sine și despre lumea din jur. Fiecare dintre noi folosește persoana întâi singular nu doar pentru a spune “asta mă doare” sau “asta îmi place”, dar și pentru a-și formula intențiile și pentru a le duce la îndeplinire prin acțiune. *Ego, ergo ago*. Noi suntem în esență agenți, nu doar ființe simțitoare – într-adevăr, nu am putea fi ființe simțitoare dacă nu am fi capabili să reacționăm adecvat la plăceri și dureri. Mill, autorul cărții *Despre libertate*, trebuia să observe că fiecare persoană este, pentru sine, prima persoană, și era necesar să respecte această plasare la persoana întâi. Dar supunerea filială i-a interzis orice încercare de a abandona explicit utilitarismul. El a continuat să fie utilitarist în proprii săi ochi susținând că utilitarismul a fost greșit înțeles de către critici și că, așa cum l-a expus el, e scutit de toate imperfecțiunile pe care aceștia i le-au atribuit. Aceasta a presupus o acrobație intelectuală surprinzătoare pe care alții, cum suntem noi, o vor explora și exploata. Totuși, ceea ce vreau eu să construiesc e un cadru mai larg de interpretare, bazat pe teoria jocurilor, care va reuni în același context cele două jumătăți ale gândirii lui Mill și prin intermediul căruia vom vedea cum sunt legate acestea între ele, cu plusurile și cu minusurile lor.

Teoria jocurilor ne ajută să înțelegem felul în care raționăm atunci când luăm decizii ce implică mai mult de o persoană. Ea ne arată de ce trebuie să țin cont de deciziile altora, ca și de propriile mele decizii, de ce ceea ce s-a petrecut în trecut e relevant în aceeași măsură ca și ceea ce s-ar putea întâmpla în viitor și de ce valorile mele trebuie să se dezvolte astfel încât să acopere binele comun și nu doar binele meu individual.

În teoria jocurilor fiecare decident, sau “jucător”, are un număr de opțiuni care pot da naștere unei serii considerabile de “rezultate”, în funcție de alegerile făcute de el sau de ceilalți jucători. Astfel, dacă sunt patru jucători, fiecare având trei posibilități de a acționa, vor rezulta 81 ($3 \times 3 \times 3 \times 3$) de rezultate posibile. Fiecare rezultat este evaluat de către fiecare dintre jucători potrivit propriului său sistem de valori iar valoarea pe care jucătorul i-o atribuie e numită “câștigul” său. Câștigul este în mod obișnuit exprimat în termeni numerici, cu observația că avem de-a face aici cu utilitățile cardinale, interpersonale, în care cred utilitariștii, deși nu este nevoie să presupunem că aceste valori sunt întotdeauna cardinale și interpersonale; în majoritatea cazurilor este suficient ca fiecare jucător să-și poată stabili ordinea

priorităților între diferitele rezultate care ar putea decurge din alegerile lui și ale celorlalți².

În cadrul teoriei jocurilor, utilitarismul acțional este cazul limită al unui joc cu un singur jucător, în care utilitaristul este singurul decident care hotărăște să producă acel rezultat care va avea câștigul cel mai mare. Utilitarismul acțional manifestă ceea ce Bernard Williams a numit o atitudine "gubernamentală" (*Government House attitude*): e binevoitor, vrea să facă ceea ce este cel mai bine pentru poporul său, dar nu ține cont de faptul că acțiunile oamenilor, în conformitate cu propriile lor convingeri, reprezintă ceva ce trebuie tratat cu respect. E binevoitor, dar e vorba de o bunăvoință despotică.

Odată ce recunoaștem că există și alți oameni care iau propriile lor decizii, pot apărea conflictele, unele dintre ele cu rezultate care ne pun în încurcătură. *Dilema prizonierului* este cel mai cunoscut și mai familiar dintre ele. Ea a fost sesizată pentru prima dată de către Protagoras³, l-a impresionat profund pe Platon⁴ și mai târziu pe Hobbes care a făcut din ea temelia argumentului său pentru Leviathan. Forma sa modernă i se datorează lui A.W. Tucker⁵.

Jocul 1: *Dilema prizonierului*

În matricea de pe pagina următoare reprezentăm doi decidenți, eu în partea stângă și un altul în partea de sus. Fiecare dintre noi are două opțiuni ce pot produce patru rezultate posibile, fiecare având propria valoare – sau câștig – pentru mine (valoare arătată de numărul aflat în partea din stânga-jos a rezultatului), respectiv pentru celălalt individ (arătat de numărul aflat în partea din dreapta-sus a rezultatului).

² Această abordare este luată din unele expuneri mai ample ce pot fi găsite în M. R. Griffiths and J. R. Lucas, *Ethical Economics*, Macmillan, Basingstoke, 1996, Appendix A, pp.222-229; J.R. Lucas, *On Justice*, Clarendon Press, Oxford, 1980, cap. 3, pp. 35-71; J.R. Lucas, *Responsibility*, Clarendon Press, Oxford, 1993, cap. 4. 4.7, pp.69-72.

³ Platon, *Protagoras*, 322ff

⁴ Platon, *Republica II*, 369-171

⁵ Formularea lui Tucker nu a fost expusă într-un studiu, ci într-o sală de curs. Așa cum a scris S.J. Hagenmayer în *The Philadelphia Inquirer* ("Albert W. Tucker, 89, Famed Mathematician", Thursday, Feb. 2, 1995, p. B7), "În 1950, pe când se adresa unei audiențe formate din psihologi la Universitatea Stanford unde era profesor invitat, dl. Tucker a creat *Dilema prizonierului* pentru a ilustra dificultățile ce apar în încercarea de a analiza" anumite tipuri de jocuri. "Explicația simplă a dlui. Tucker a fost de atunci sursa a numeroase articole în domenii diverse precum filosofie, etică, biologie, sociologie, științe politice, economie și, bineînțeles, teoria jocurilor."

		Celălalt	
		îmi face un rău	nu-mi face un rău
Eu	îi fac un rău	<p style="text-align: right;">1</p> <p>Viața amândurora este grea, brutală și scurtă, dar cel puțin ocazional reușesc să iau ceva din bunurile celuiilalt.</p> <p>1</p>	<p style="text-align: right;">0</p> <p>Viața mea este încântătoare. Mă bucur de asigurarea că el nu-mi face nici un rău și de libertatea de a-i face rău oricând doresc. Viața pentru el este grea, brutală și scurtă, fără nicio consolare.</p> <p>10</p>
	nu îi fac un rău	<p style="text-align: right;">10</p> <p>Viața pentru el este încântătoare. Se bucură de asigurarea că nu i se va face nici un rău combinată cu libertatea de a-mi face rău oricând dorește. Viața mea este grea, brutală și scurtă, și chiar când am ocazia de a profita de pe urma lui nu o fac.</p> <p>0</p>	<p style="text-align: right;">6</p> <p>Viața amândurora este acceptabilă, dar între anumite limite. Amândoi ne bucurăm de siguranță față de atacurile celuiilalt, dar amândoi suntem frustrați de imposibilitatea de a ne manifesta pe deplin potențialul personal. Viața este confortabilă, liniștită și lungă, dar lipsită de autenticitate; și amândoi suferim de <i>mauvaise foi</i>.</p> <p>6</p>

Ceea ce demonstrează *Dilema prizonierului* este lipsa de raționalitate a ideii de “eu înainte de toate” (*me-firstism*). Dacă rămânem doar la persoana întâi singular, vom adopta o politică aptă ca, în anumite situații, să producă rezultate mult mai rele pentru fiecare dintre noi decât am fi obținut dacă am fi trecut de la singular la plural și am fi ținut astfel cont de ceea ce era cel mai bine pentru noi toți. O ducem mai bine dacă cooperăm unii cu alții decât dacă eu, și aceasta e valabil pentru toți ceilalți, mă gândesc doar la mine însumi.

Jocul 2 : Regula drumului

		Dl. Knight	
		o ia la dreapta	o ia la stânga
Dl. Chevalier	à droite	5 Fiecare trece pe lângă celălalt în siguranță 5	0 Se ciocnesc 0
	à gauche	0 Se ciocnesc 0	5 Fiecare trece pe lângă celălalt în siguranță 5

Regula drumului arată importanța convențiilor, a “normelor de coordonare”, care permit jucătorilor dintr-un joc cu mai multe persoane să-și armonizeze împreună deciziile astfel încât să asigure producerea rezultatelor pe care le preferă cu toții. Când vine vorba de condusul mașinii, de comunicare, de dans, precum și de multe alte activități sociale, trebuie să ne coordonăm acțiunile unii cu ceilalți astfel încât să ne armonizăm eforturile și să evităm conflictele. Reprezentăm în matrice în mod schematic doi automobiliști care se apropie unul de celălalt, dl. Knight și dl. Chevalier, și care trebuie să se ferească pentru a nu se ciocni (cu câștigul domnului Knight în partea din dreapta-sus a fiecărui rezultat, și cu câștigul domnului Chevalier în partea din stânga-jos). Dacă amândoi fac dreapta sau dacă amândoi fac stânga, vor trece unul pe lângă celălalt în siguranță: esențialul este că niciunul dintre ei să nu decidă ceea ce el, de unul singur, crede că este cel mai bine, ci să respecte amândoi o anume convenție, sau regulă, sau lege, sau înțelegere reciprocă. Cu alte cuvinte, asta înseamnă să spui că nu ar trebui să încerc să fac ceea ce mi se pare mie că ar produce cele mai bune consecințe, ci ar trebui să acționez cu încredere așa cum ceilalți se așteaptă că o voi face. Ar trebui să conduc numai pe partea stângă și să nu fac depășiri imprudente, să dau prioritate atunci când celălalt șofer are prioritate, să o iau înainte atunci când am eu prioritate, astfel încât ceilalți șoferi să-și dea seama cum să se raporteze la mine și să poată să-și planuiască propriile mișcări în consecință. În absența unor reguli recunoscute în mod public e inevitabil caracterul imperfect al informațiilor despre acțiunile viitoare ale agenților liberi: o deontologie a respectării normelor este cheia succesului atunci când o simplă strategie de maximizare este imposibilă și când fiecare jucător

trebuie să țină pasul cu ceilalți prin respectarea de către toți participanții la joc a aceluiași convenții relevante. Indiferent care ar fi atracțiile aparente ale consecinționismului pentru un operator izolat, ele devin iluzorii, chiar și după strandardele consecinționiste, de îndată ce agentul se vede pe sine nu ca pe un solipsist singuratic, ci ca pe o persoană între multe altele, fiecare trebuind să îi recunoască pe ceilalți ca inițiatori ai unor acțiuni făcute cu mintea proprie și ale căror decizii pot fi anticipate doar dacă aderă la reguli bine-cunoscute.

Jocul 3: Bătălia între sexe

În *Bătălia dintre sexe*, "el" și "ea" vor să-și petreacă vacanța împreună, dar el ar prefera să meargă la munte în Alpi, pe când ea ar prefera să-și petreacă vacanța împreună la plajă, pe malul mării. Având în vedere că pentru fiecare dintre ei a doua alegere favorită este cu mult mai bună decât a treia sau a patra alternativă, ar merita să o accepte amândoi pe aceea, dacă cea mai bună variantă nu poate fi obținută. Și atunci ar merita să-l faci pe celălalt să creadă asta. Dacă ea poate face o criză de nervi și spune că sub nicio formă nu va putea suporta Alpii și că nu va merge acolo nici în ruptul capului, atunci el, dacă este un individ rezonabil, va renunța la visul unei vacanțe alpine și se va mulțumi cu marea, pentru că aceasta i-ar plăcea de două ori mai mult decât i-ar plăcea să meargă singur la munte. Dar în egală măsură el ar putea crede că a sosit momentul să se poarte bărbătește și să o lase pe biata femeie să înfrunte situația: fie merge unde vrea el, fie se duce singură. Și atunci când va fi clar că acestea sunt singurele variante pe care le are, ea nu va avea de ales și va trebui să cedeze, adică să-și cumpere un rucsac în locul unui nou costum de baie. Este, așadar, irațional să te ghidezi doar după câștigul atașat rezultatelor disponibile la un moment dat, pentru că astfel se creează posibilitatea ca celălalt să-ți manipuleze opțiunile. Dacă e să-mi păstrez autonomia, nu pot fi pur și simplu un consecinționist direct. Odată ce tu știi că eu mă ghidez doar după consecințe, mă poți influența să fac orice dorești dacă aranjezi situația în așa fel încât, atunci când vine timpul să iau o decizie, rezultatul disponibil cel mai puțin rău pentru mine va fi acela care se conformează planurilor tale. Raționalitatea ne cere, în schimb, să ne extindem aria lucrurilor avute în vedere atât ca timp cât și ca persoane.

Ea			
		merge în Alpi	merge la mare
El	merge în Alpi	8 Încântător pentru el, bine pentru ea 10	4 "Aș fi vrut să fii și tu aici" 4
	merge la mare	0 Deplorabil pentru el; deplorabil pentru ea. 0	10 Bine pentru el; încântător pentru ea. 8

Este deseori un avantaj să fii în stare să te constrângi în mod absolut sau, echivalent, să elimini unele opțiuni în mod absolut. Strategia *Distrugerii reciproc-garantate* funcționa doar dacă fiecare parte credea că cealaltă nu se ghidează exclusiv după considerente de tip consecinționist și că, dacă ar fi fost atacată, ar fi răspuns efectiv, chiar dacă în acel moment nu ar fi avut niciun avantaj comportându-se astfel. Pentru a consolida această așteptare, au fost construite dispozitive mecanice care să funcționeze în mod automat în eventualitatea unui atac nuclear, fără posibilitatea de a fi închise de vreun supraviețuitor consecinționist. Într-o manieră mai puțin sumbră, întreaga logică a facerii și a ținerii promisiunilor este să ne asigure că anumite acțiuni ale unui agent nu trebuie modificate pur și simplu pe temeiul existenței unor factori care au fost viitori dar care, o dată cu trecerea timpului, au devenit trecuți. Dacă ignorăm toate considerentele noastre trecute nu numai că devenim vulnerabili în fața manipulării, dar oferim doar o explicație parțială a contextului în care sunt luate deciziile noastre și din care își dobândesc ele semnificația. Nu pot fi orientat în mod coerent doar înspre viitor odată ce recunosc că viitorul meu va deveni la un moment dat trecut.

Fiecare dintre aceste argumente este de tipul *reductio ad absurdum*. Începem prin a presupune, așa cum făceau și economiștii clasici, că raționalitatea poate fi definită în termeni de maximizare a câștigurilor viitoare și apoi arătăm că o astfel de definiție este auto-contradictorie în propriii săi termeni. *Dilema prizonierului* ne arată că acesta ar trebui să țină cont nu doar de existența, ci și de interesele și idealurile celorlalți oameni, și că este irațional să ignori punctul de vedere al colectivității. *Regula drumului* arată că este mult mai bine să respecti regulile decât să încerci, așa cum te-ar sfătui utilitariștii acționali, să săvârșești acțiunea care va avea cele mai bune

consecințe: fiecare dintre noi trebuie să admită că nu este singur pe lume, că nu ține de el să decidă ce curs al evenimentelor va avea loc, și că adesea cel mai bun lucru pe care-l poate face este să se armonizeze cu ceea ce e cel mai probabil că vor face și ceilalți. *Bătălia între sexe* ne arată că este irațional să te gândești doar la rezultatele viitoare; agentul are atât un trecut cât și un viitor și ar trebui să hotărască ceea ce face ținând cont atât de deciziile luate în trecut cât și de ceea ce va urma în viitor. Contrar punctului de vedere static, solipsist, orientat doar spre viitor și în mod exclusiv individualist al economiștilor clasici, suntem nevoiți, reflectând la aceste trei cazuri, să recunoaștem că raționalitatea este dinamică și că ne determină să adoptăm un punct de vedere personal mai vast și mai extins din punct de vedere temporal cu privire la ce anume e conținut în deciziile pe care trebuie să le luăm.

Mill ar fi fost încântat de această abordare teoretică din perspectiva teoriei jocurilor chiar dacă argumentele expuse aici nu i-ar fi oferit tot ceea ce l-ar fi interesat – dreptatea, în particular, nu poate fi salvată doar pe baza acestor argumente, ci are nevoie de argumente suplimentare în care să ne adresăm la persoana a doua celor împotriva cărora se iau decizii nefavorabile⁶. Cu toate acestea, abordarea din perspectiva teoriei jocurilor ar fi repus în drepturi individul în calitatea sa de agent având propriile planuri și proiecte și nefiind numai o unitate capabilă doar să simtă durerea și plăcerea; ea ar fi oferit, de asemenea, o cale de a susține cu argumente concluzia, pornind de la premisa că fiecare își dorește propria fericire, că noi toți trebuie să urmărim fericirea tuturor. Totuși, Mill, ca un utilitarist zelos ce era, ar fi continuat să fie îngrijorat gândindu-se cum să-și convingă tatăl să conjuge. Dar principiul conjugării era deja acceptat prin aceea că utilitarismul acțional ținea cont de consecințele viitoare, nu doar de cele prezente. Principiul originar al plăcerii ne îndemna să urmărim plăcerile prezente – *to paron hedu* – și abia după multă educație și multă săcâială am putea trece de la hedonismul simplu la prudență. Odată ce lucrurile sunt reasezate în acest fel, nu mai are noimă să obiectăm împotriva luării în considerare a altor momente de timp și a altor persoane.

Teoria jocurilor dă o mai bună explicație decidentului individual care nu ne mai apare acum ca un eu izolat, cu o viziune trunchiată asupra timpului, preocupat doar de viitor, ci ca un agent care are un trecut și care are motive să se identifice cu noi și cu ei, cu tine. Dar trebuie să ne ferim de un fel de exces de conjugare ignorând în totalitate importanța persoanei întâi singular.

⁶ Vezi pe larg, J.R. Lucas, *On Justice*, Clarendon Press, Oxford, 1980, cap. 1; sau "*The Concept of justice*", URL=<"http://users.ox.ac.uk/~jrlucas/libeqsor/justice.html">

Sfaturile prudentiale ne pot duce la ipotecarea prezentului în favoarea viitorului și o știm prea bine că adesea, în secolul al XX-lea, interesul individului a fost subordonat interesului imaginar al unei pseudo-colectivități. Secolul al XX-lea fost unul jalnic din cauză că s-a considerat că ideea de auto-determinare e mai importantă decât grija pentru indivizii reali, securitatea și libertatea lor. Motto-ul era "E mai bine să avem auto-guvernare decât guvernare bună". Îmi aduc aminte că, acum mai bine de cincizeci de ani, aveam, ca student, o dispută cu Bernard Williams la Balliol College despre Burma (Birmania), care era pe cale să își obțină independența și să fie eliberată din cătușele Imperiului Britanic. Eram sceptic și spuneam că dacă aș fi un țăran birmanez aș prefera să fiu condus de un funcționar britanic care, poate că e distant și nu se implică, dar e imparțial și judecă drept lucrurile, eventual chiar cu o anumită bunăvoință, decât să fiu condus de un bandit local, lord al războiului, care vorbește într-adevăr aceeași limbă cu mine și are aceeași culoare a pielii, dar n-are nicio remușcare în a mă oprima și asupra. Bernard nu era deloc de acord cu mine. El era în ton cu spiritul vremii. Dar acum, cincizeci de ani mai târziu, pe măsură ce ies la iveală informații despre starea dezastruoasă a acestui popor sub conducerea actuală⁷, îndoielile mele sunt, din păcate, confirmate. Dacă Mill ar fi trăit astăzi ar fi conjugat: dar n-ar fi pierdut niciodată din vedere importanța excepțională a persoanei întâi singular la timpul prezent.

Traducere de Livia-Ana Tinca și Valentin Mureșan

⁷ Vezi, de exemplu, Pascal Khoo Thwe, *From the Land of Green Ghosts: A Burmese Odyssey*, Harper Perennial, Londra, 2004.