

Minte și limbaj în context evoluționist*

Acad. Mircea Dumitru

Vicepreședinte al Academiei Române

Rezumat

Modelul științei darwiniene al evoluției indivizilor dintr-o specie prin intermediul mecanismului variațiilor interne aleatorii ale caracteristicilor biologice ale acestora, care sunt filtrate prin selecția naturală impusă din exterior de către mediul la care organismul în cauză se adaptează, a cunoscut în anii din urmă o aplicare la caracteristici non-biologice ale indivizilor și, prin urmare, o extindere semnificativă la domenii pentru care acest model nu a fost conceput în mod inițial și, așadar, la care nici nu a fost aplicat vreodată. În studiul de față explorez o astfel de aplicație și extindere la conceptul central pentru domeniul filosofiei minții și respectiv al filosofiei limbajului, anume la conceptul de *semnificație* (*meaning*).

Introducere.

Concepte, probleme, întrebări

Concepțiile dominante în filosofia actuală a minții, dezvoltate pentru a răspunde la problema minte-corp și la chestiunea caracterului fenomenal-subiectiv al conștiinței, sunt motivate de viziuni metafizice ample, de natură materialist-fizicalistă (cum sunt, de pildă, teoriile reduționiste ale identității minte-corp, sau teoriile eliminativiste). Dar această viziune materialist-reduționistă sau eliminativistă ne pune în fața unei încurcături conceptuale greu de rezolvat, anume cum să armonizăm sau să compatibilizăm proprietatea actelor noastre mentale de a fi semnificative cu procesele neurofiziologice care susțin, sau în care rezidă, aceste acte mentale.

Problema crucială pentru înțelegerea unitară a locului nostru în schema generală a lucrurilor este cum anume să dăm seama de caracterul semnificativ al proceselor mentale, dar și de conținutul produselor sau al rezultatelor la care conduc aceste procese, în contextul unor explicații naturalist-fizicaliste ale mecanismelor care susțin desfășurarea acelor procese mentale? Cum anume se integrează mintea și semnificația stărilor mentale și a actelor de limbaj în schema generală a lumii naturale? Altfel spus, cum anume este posibil sensul sau semnificația reprezentărilor mentale și ale limbajului într-o lume guvernată – și totodată închisă față – de legile fizice naturale?

Ne aflăm, fără îndoială, într-un impas și nu avem nici o idee clară cum anume să ieșim din această fundătură teoretică. Imaginea noastră globală despre natură, dar și despre actele semnificative ale minții umane (reprezentările mentale și limbajul), este profund divizată. Pe de o parte, științele care furnizează cele mai reușite și plauzibile explicații despre natura biofizică a omului conturează un tablou naturalist-fizicalist și reduționist în care, metodologic vorbind, omul este abordat și explicat aidoma unei mașinării bio-chimice complexe, guvernată de legi care aparțin domeniului închis cauzal al realității fizice și care, tocmai de aceea, poate fi obiectul unei cercetări științifice obiective care aspiră să producă explicații cauzale.

Pe de altă parte, însă, psihologia simțului comun, sau pur și simplu imaginea noastră manifestă obișnuită despre om, îl reprezintă pe acesta ca pe un agent rațional și conștient, ale cărui decizii și acțiuni fac obiectul unor explicații intenționale

* Comunicare în cadrul manifestării „Principiul divergenței și cel al selecției naturale la Charles Darwin”, masă rotundă organizată de Secția de filosofie, teologie, psihologie și pedagogie a Academiei Române, în Clubul Academicienilor, la data de 29 martie 2024.

sau teleologice bazate pe opinii, preferințe și scopuri: stările mentale (atitudinile propoziționale) au un conținut reprezentational, iar oamenii acționează pe baza acestor reprezentări după tiparul „silogismului” practic dorințe/preferințe – opinii – acțiuni; oamenii *doresc* anumite lucruri sau *urmăresc* să se producă anumite stări, respectiv *preferă* o stare altei stări, au *opinii* (reprezentări mentale) despre ce acțiuni trebuie întreprinse pentru a realiza dorințele respective și *decid* cum anume să canalizeze energia voinței pentru a acționa în vederea satisfacerii scopurilor lor de a-și îndeplini dorințele sau preferințele.

Din păcate, luate ca atare, imaginea științific-reducționistă, pe de o parte, și imaginea manifestă a simțului comun, pe de altă parte, nu par să fie ușor de integrat și de compatibilizat. Ele au, însă, luate separat și fiecare în contextul ei de validitate și de legitimare, puternice temeiuri care le conferă o plauzibilitate ridicată și nu există nici o motivație mai profundă, teoretică sau practică, pentru a renunța la una dintre ele. Dar atunci, cum anume ar putea fi menținute ambele imagini?

Așa cum este de așteptat, atunci când avem de-a face cu un astfel de impas teoretic, soluțiile la această problemă a integrării, care sunt documentate în istoria filosofiei și a științei, au pendulat între cele două poziții extreme ale spectrului răspunsurilor logic-posibile și acceptabile la problema minte-corp, anume dualismul și respectiv reducționismul sau eliminativismul materialist. Răspunsurile se grupează în două categorii principale de reacții la întrebarea dacă poate fi realizată integrarea explicativ-teoretică a minții și a semnificației reprezentărilor mentale și a limbajului în domeniul atotcuprinzător al lumii fizice.

Starea actuală a problemei poate fi caracterizată, în mare, ca o confruntare argumentativă între două poziții metafizice radical opuse: (i) *incompatibilismul* și (ii) *compatibilismul naturalist*.

Astfel, pe de o parte, există grupul de concepții care se subordonează unui concept *incompatibilist*, potrivit căruia perspectiva descrierii științifice, obiective, asupra fenomenelor mentale-conștiente și punctul de vedere manifest, al simțului comun, asupra acelorași fenomene nu pot fi armonizate și nu sunt coerente luate împreună. Din această categorie fac parte două grupe de concepții

care sunt diametral opuse, dar care sunt elaborate pe baza aceluiași principiu de excluziune: una dintre perspective o anulează pe cealaltă; una face apel la intuițiile noastre comune cu privire la dimensiunea conștientă a mentalului și socotește că perspectiva obiectiv-științifică nu este aptă să cuprindă și o explicație a domeniului mentalului, în timp ce cealaltă consideră viziunea simțului comun ca fiind esențialmente greșită și că numai știința poate oferi o abordare corectă a fenomenelor psihologice și subiectiv-conștiente. Pe scurt, pentru a pune nume acestor perspective incompatibiliste, avem în vedere aici, pe de o parte, concepția carteziană, potrivit căreia, în fondul lucrurilor, perspectiva obiectiv-științifică este iremediabil incompletă pentru că nu ne poate oferi o explicație a conștienței ființelor umane și pe de altă parte, concepția materialist-eliminativistă, pentru care, dimpotrivă, psihologia simțului comun este o teorie falsă, care trebuie respinsă și eliminată de o (viitoare) abordare științifică obiectivă, completă și corectă.

Filosofia contemporană a minții este dominată, totuși, de un grup de concepții care se plasează într-un perimetru conceptual și argumentativ *compatibilist-naturalist*. Astfel, concepția standard în acest domeniu foarte activ al filosofiei contemporane este funcționalismul. Conform acestei doctrine, psihologia simțului comun este o teorie generală despre arhitectura, structura și organizarea cogniției umane, care definește într-un mod implicit și naiv natura stărilor mental-cognitive umane, cum ar fi, în principal, opiniile, preferințele, dorințele, sau emoțiile. Toate aceste stări sunt concepute ca niște stări funcționale. Opiniile, preferințele și toate celelalte stări mentale sunt genuri sau tipuri funcționale definite prin intermediul rolului lor cauzal, adică prin ce anume fac sau produc ele în economia arhitecturii mentalului, și nu prin constituția lor fizică, adică prin natura sau morfologia substratului material pe care se produc aceste stări mentale. În principiu, o stare mentală, definită suficient de abstract și prin apel la țesătura de relații cauzale în care este prinsă, fie drept cauză a altor stări sau acțiuni, fie drept efect al altor stări sau acțiuni, poate fi realizată pe substraturi sau suporturi materiale diverse, a căror morfologie sau structură poate varia mult de la o

substanță la alta. Această caracteristică a uneia și aceleiași stări-tip mentale de a fi instanțiabilă pe suporturi fizice diferite poartă numele de *realizabilitate multiplă*. Acesta este conceptul central folosit de funcționaliști pentru a submina argumentarea materialist-reducționistă sau teoria identității-tip minte-creier.

Partea bună a lucrurilor cu această viziune compatibilist-naturalistă este aceea că funcționalismul oferă perspectiva unei soluții la problema integrării: oamenii sunt deopotrivă structuri biochimice extrem de complexe și agenți raționali, intenționali și conștienți. Versiunea cea mai elaborată, frecventată și influentă a funcționalismului actual este așa-numita teorie computațională a minții, care oferă drept concept explicativ integrator ideea că rolul unei stări mentale de a media între stimuli și comportament trebuie să fie explicat în termeni computaționali: stările mentale sunt computații sui-generis asupra reprezentărilor mentale.

Oricât de puternică și fertilă, concepția funcționalistă trebuie să se confrunte, totuși, cu serioase dificultăți conceptuale și explicative. Astfel, pe de o parte, incompatibiliștii de factură dualistă insistă asupra ideii că funcționalismul oferă prea puțin, căci această concepție nu ne oferă o explicație serioasă și credibilă a relației dintre cogniție și conștiința subiectivă; pe de altă parte, incompatibiliștii eliminatiști se plâng de faptul că funcționalismul definește arhitectura cauzală a stărilor mentale la un nivel mult prea abstract, nespecific și generic, în timp ce cogniția reală este un proces complex mult mai înrudit în structura neuro-fiziologică a creierului.

Este perfect legitim, atunci, ca într-o astfel de situație teoretică și conceptuală închisă ermetic să intervină în mod firesc nevoia și tentația teoreticianului de a explora și de a experimenta diverse ipoteze explicative și hermeneutice. Una dintre aceste ipoteze, care și-a probat fertilitatea în câteva domenii teoretice plasate dincolo de domeniul originar și original al biologiei în care a fost pentru prima oară formulată, este ipoteza evoluționistă darwiniană. Ajungem, în felul acesta, la ideea de a aplica temele și conceptele darwiniene (evoluționiste) în filosofia minții și a limbajului.

Darwinianismul ca soluție compatibilist-naturalistă pentru problema integrării

Întrebarea care ne preocupă aici va fi, așadar, aceasta: este darwinianismul o soluție compatibilist-naturalistă pentru problema integrării? Și dacă da, atunci care sunt promisiunile teoretice ale acestei abordări și care sunt, complementar, dificultățile cu care va trebui să ne confruntăm?

Angajarea față de un răspuns sau altul la întrebările precedente ne cere să lămurim, în mod prealabil, o presuposiție deopotrivă teoretică și metodologică pe care aceste interogații o poartă cu sine. Anume, aceea că există o pretenție epistemică substanțială legitimă că modelul darwinian poate fi exportat din biologie în alte domenii ale cunoașterii, inclusiv la câteva domenii din științele sociale și umaniste. Darwinianismul s-ar bucura de acest gen de privilegiu, deoarece, în forma sa abstract-structurală, explicația darwiniană a evoluției prin mecanismul selecției impuse de ecologia externă asupra variațiilor produse de natura internă a entităților supuse evoluției ar avea efectul liberator (devastator?) al unui „acid universal” (Daniel Dennett), care ne învață că modul specific în care poate să se producă / să apară complexitatea este un proces a cărui sintaxă sau alcătuire abstract-formală este neutră față de domeniul în care se produc acele procese evolutive. Mulți naturaliști lucrează azi, de fapt, pe baza acestei asumptii. Ei dezvoltă modelul prin articularea și aplicarea lui la alte domenii decât acela avut în vedere, în mod originar, de către autorul său principal.

Redus la forma sa cea mai abstractă și generică, modelul explicativ darwinian rezidă în următorul mecanism de filtrare exogenă sau externă impus asupra unor variații aleatorii endogene sau interne: organismul, entitatea, sau unitatea care trece printr-un proces evoluționist produce, prin intermediul unor mecanisme endogene, *variații aleatorii* ale unor trăsături definitorii pentru entitatea respectivă; mediul natural, social, conceptual etc. *selectează* doar acele entități ale căror variații aleatorii se dovedesc a avea un potențial adaptativ ridicat și le selectează pe acestea în virtutea acelor trăsături; dat fiind un anumit interval de timp rezonabil, selecția contribuie, în manieră non-finalistă și non-intențională, la fixarea acelor trăsături cu

potențial adaptativ ridicat prin intermediul mecanismului transmiterii acestora la generațiile care au trecut testul luptei pentru supraviețuirea celui mai apt. Heritabilitatea trăsăturilor cu potențial adaptativ ridicat generează schimbări adaptative locale, indiferent de natura entităților: gene, organisme, idei etc.

Dar ce anume urmează să fie abordat prin prisma acestui model explicativ care aspiră să aibă vocație universală? Cum am văzut, deja, raționalizarea sau explicarea acțiunii umane în teoria interpretării (sau psihologia simțului comun) apelează la conexiunea pe care o asigură un gen special de silogism practic sau acțional: oamenii acționează pe baza unor stări mentale individualizate în termenii conținuturilor lor (opinii, dorințe, preferințe etc.). Modelul simplificat al acestui silogism este structura tripartită *Dorințe / Preferințe – Opinii (Credințe) – Acțiuni*, prescurtat D/P – O – A.

Stările mentale sunt, în mare vorbind, reprezentări ale lumii care au un conținut intențional (sunt *despre* ceva care există în mod real sau ideal). Marea problemă, în care s-au împotmolit toate încercările sistematice dezvoltate recent în filosofia minții de orientare analitică, este aceasta: Ce anume este o reprezentare? Această întrebare fundamentală se ramifică în alte câteva chestiuni la fel de dificile și de urgente. Ce face ca substanța sau conținutul unei reprezentări să fie exact ceea ce este și, mai ales, cum se face că în marea majoritate a cazurilor, în mod tipic, atunci când avem nevoie de un anumit conținut pentru a efectua o anumită acțiune, acest conținut devine disponibil, din punct de vedere epistemic, conștiinței noastre? Adică, plasând întrebările la un nivel mai specific, ce face, de pildă, ca reprezentarea mea (conceptul meu) de, să zicem, *hipopotam* să fie despre hipopotami? Ce face ca anumite structuri neuro-chimice din creier să fie constitutive pentru (sau identice cu?) un *simbol* al hipopotamilor?

Dar întrebarea din urmă pare să creeze o disonanță în setul de întrebări, destul de omogene, care îi premerge. Căci punerea ca atare a problemei relațiilor dintre structurile neuro-chimice din creier și reprezentări mentale, concepte sau simboluri pare să ne transpună într-un alt registru de cunoaștere. Relațiile naturale dintre ceea ce se întâmplă în creier și reprezentările noastre mentale sunt mai

degrabă relații cauzale neuro-fiziologice decât relații logice, criteriale sau convenționale, adică acel gen de relații care te aștepti să dea seamă de natura reprezentățională a unui concept sau reprezentări. Cauzalitatea naturală pare să aparțină altei familii de concepte și să determine un alt gen de cercetare decât acela cu care suntem obișnuiți în cercetările tradiționale de filosofie a limbajului și a minții întreprinse cu mijloace logice și analitic-conceptuale.

Naturalistul substituie relația convențională cu una fizic-cauzală. Dar atunci, nu este câtuși de puțin surprinzătoare întrebarea acestuia cu privire la elucidarea naturii semnificative a reprezentărilor în termenii unor conexiuni naturale. Modul său preferat de a pune această problemă va fi acesta: *Ce relație naturală există între un simbol mintal și referentul său din lumea externă? Ce conferă semnificație simbolurilor mintale?*

Dar simplitatea formulării întrebării nu trebuie să ne înșele cu privire la dificultatea sarcinii ca atare de a elucida această relație dintre reprezentarea mintală și ceea ce este reprezentat. De fapt, această problemă este un adevărat „coșmar” pentru naturalist și pentru întregul său program de naturalizare a semanticii, deoarece conceptele nu sunt nici imagini, nici modele sau prototipuri prezente în creier ale lucrurilor, nici abilități de recunoaștere.

Exact în legătură cu această dificultate ne punem întrebarea dacă este util modelul darwinian „Variabilitate – Selecție – Heritabilitate” pentru sarcina de a explica conținutul mental și semanticitatea reprezentărilor mentale. Prima sugestie serioasă care poate să pună în mișcare o strategie naturalistă de tip darwinian pentru explicarea naturii reprezentării este aceea că un concept al meu de *hipopotam* este despre hipopotami datorită **istoriei naturale a acestui concept**. În acest context explicativ, un concept este un mecanism adaptativ – el direcționează comportamentul meu în relație cu hipopotamii. Mecanismele mentale sunt adaptări.

Elaborarea acestei idei conduce, așa cum se poate observa de pildă în argumentele construite de Fred Dretske și respectiv de Jerry Fodor, la susținerea că simbolul mintal co-variază, în mod tipic, cu ceea ce desemnează sau semnifică acesta.

Dar, oricât de plauzibilă și de promițătoare pare să fie această rută explicativă, ideea nu este lipsită de dificultăți și de fapt ea trebuie să facă față unor probleme serioase. Iată câteva dintre acestea: sunt situații în care co-variația avută în vedere aici lipsește, sau în care co-variația este instabilă, sau, de asemenea, avem situațiile în care conceptele au în mod evident caracter reprezentational, dar nu au cu ce anume să co-varieze, deoarece conceptele în cauză sunt vide, iar o astfel de situație este foarte greu de acomodat cu genul de abordare semantică naturalistă pe care-l urmărim aici.

O soluție la dificultățile care se ridică în fața modelului schițat mai sus este modelul alternativ teleosemantic, datorat lui David Papineau și respectiv lui Ruth Millikan, bazat pe conceptul de funcție biologică. Potrivit acestui nou model, gândurile/conceptele au funcții biologice: ele orientează comportamentul prin care organismul se adaptează la trăsături ale mediului. Evoluția, socotește Ruth Millikan, este aceea care modelează și formează creierul nostru să producă reprezentări prin intermediul anumitor metode de interacțiune fizică cu ceea ce este reprezentat, interacțiuni care sunt controlate de către reguli foarte abstracte de proiecție. Ceea ce conectează o percepție sau un gând la o stare de lucruri pe care o reprezintă este un anumit gen de *desing* evoluționist (cf. Ruth Millikan, 1993). Gândul lui *Y* este *despre X* deoarece funcția gândului este de a-i permite individului *Y* să se adapteze la prezența lui *X* în mediul său. Eroarea de reprezentare se datorează eșecului funcției, tot așa cum o malfuncție a unui organ rezidă în eșecul funcției sale proprii biologice.

Schematic, planul de naturalizare a reprezentărilor mentale și a celor semantice se articulează în jurul următoarei idei, pe care Ruth Millikan o exprimă foarte limpede: „Sistemele cognitive sunt proiectate de către evoluție pentru a produce imagini abstracte ale mediului înconjurător al organismului și pentru a fi orientate de către aceste imagini în producerea acțiunilor potrivite. Aceasta poate să implice derivarea unor imagini adiționale prin combinarea unor imagini vechi în acord cu anumite principii (inferențe, raționamente). Poate să implice, de asemenea, formarea reprezentărilor scopurilor, imagini care ghidează organismul să producă sau să evite ceea ce reprezintă (a-și ima-

gina, a se teme că, a dori, a avea intenția). În cazul oamenilor, *designul* biologic include principii foarte generale în concordanță cu care sunt formate conceptele pe baza inputului senzorial corespunzător, inclusiv inputul din percepția pe care o mijlocește limbajul. Formarea conceptelor adecvate este învățarea de a reprezenta sau de a cartografia în gândire ceea ce *este* identic *ca fiind* identic. Pe această abilitate sunt construite abilitățile de a face judecăți, apoi inferențe corespunzătoare și, de asemenea, abilitățile de a dobândi dorințe potrivite și de a le satisface.” (Ruth Millikan, 1993, pg. 11.)

Dificultăți și probleme pentru modelul darwinian-evoluționist

Naturalismul semantic este un program foarte frecventat azi. Cred, însă, că el se confruntă cu dificultăți serioase, greu de surmontat. Desigur, pe de o parte, semantica cuvintelor (word meaning), bazată pe conceptul de intenție comunicativă (modelul Grice), este mai ușor de acomodat evoluționismului. Dar, pe de altă parte, semantica de adâncime a structurilor propoziționale, teoretizată și elucidată fie sub o formă veri-condițională (Tarski, Davidson), fie sub una model-teoretică (Tarski, Montague), pare să fie mult mai greu, dacă nu chiar imposibil de acomodat!

Problema cea mai dificilă aici este înțelegerea în manieră evoluționistă a structurii sintactice de adâncime a limbajului. Dar nu numai! Dacă adoptăm drept fundal și prototip al psihologiei cognitive teoria computațională a minții, în care gândirea este considerată un calcul sui-generis cu reprezentări mentale și concepte, iar conceptele / gândurile sunt structuri de informații, atunci ce genuri de dovezi avem că aceste structuri au cunoscut o evoluție semnificativă într-o perioadă istorică rezonabilă? Pare să fie cât se poate de clar și de evident că explicațiile pe care oamenii le dau comportamentelor propriilor lor semeni în termenii conținuturilor atribuirilor de atitudini propoziționale nu s-au modificat semnificativ în lungi perioade de timp. (*X* l-a denunțat pe *Y* pentru că a crezut că *Y* l-a trădat și s-a temut că *Y* urmărește să-i facă rău etc. etc.)

Revenind la limbaj, însă, trebuie remarcat că teoria evoluționistă pune la grea încercare imaginea ineistă despre sintaxă și, în general, despre limbaj,

care se desprinde din concepția, standard astăzi, a lui N. Chomsky. Este foarte greu de armonizat, pe de o parte, o concepție care susține că limbajul este o specializare cognitivă distinctiv umană, că învățarea unei anumite limbi rezidă în învățarea unui set de reguli abstracte, iar cunoașterea noastră a acestor reguli este codificată fizic în creier și că dispunem de informații bogate înnăscute despre forma generală a limbajului, cu concepția evoluționistă, pe de altă parte, potrivit căreia limbajul este o adaptare. Dacă lucrurile ar sta așa, atunci limbajul ar trebui să cunoască o pregnantă evoluție, dar evoluția limbajului, mai ales la nivelurile profunde ale sintaxei, nu este un fenomen evident.

Nu este mai puțin interesant să reflectăm la tema epistemologică a profilului unei concepții evoluționiste despre conținut mintal și limbaj: la ce gen de teorie a conținutului, de semantică și de epistemologie conduce modelul V-S-H? Modelul are un profil mixt. El combină trăsături și elemente care corespund unor paradigme distincte ale teoretizării limbajului și mentalului. Astfel, dimensiunile variabilității și heritabilității din modelul evoluționist corespund bine profilului epistemologic al teoriei computaționale a minții și sunt compatibile cu teoriile internaliste despre conținutul mental și semantic. Caracteristica variabilității interne a trăsăturilor filtrate prin selecție este compatibilă cu cerința semantică internalistă de a identifica și caracteriza conținutul atitudinilor propoziționale (al opiniilor, mai ales), fără apel la factori externi. Pe de altă parte, dimensiunea selecției naturale converge cu externalismul conținutului mental și semantic, iar din punct de vedere epistemic se integrează programul unei epistemologii reliabiliste (*epistemic reliabilism*).

La prima vedere, problemele acestui model hibrid mixt sunt problemele care se acumulează cu fiecare dintre cele două mari genuri de programe de cercetare ale semanticității mentalului și limbajului, adică problemele externalismului, foarte acut puse în relief de criticile de genul acelor

elaborate de J. Fodor, dar și problemele internalismului, evidențiate de criticile unor generații întregi de gânditori care merg, de pildă, pe urmele lui H. Putnam.

O sugestie pentru o posibilă depășire a acestui impas în care suntem prinși de concepțiile care au polarizat câmpul cercetărilor în domeniul teoriei conținutului și ale interpretării este aceea ca filosofii minții și ai limbajului să lucreze, prin analogie cu teoria sintetică a evoluției, în vederea construirii unei teorii sintetice a conținutului mintal și a semnificației limbajului.

Toate aceste sugestii, doar schițate la sfârșitul acestui studiu, sper că vor fi dezvoltate și completate într-un studiu viitor.

Bibliografie

Fred Dretske, *Explaining Behavior: Reasons in a World of Causes*, A Bradford Book, The MIT Press, 1991.

Fred Dretske, *Naturalizing the Mind*, The MIT Press, 1997.

Jerry Fodor, *Psychosemantics: The Problem of Meaning in the Philosophy of Mind*, A Bradford Book, The MIT Press, 1989.

Jerry Fodor, *A Theory of Content and Other Essays*, A Bradford Book, The MIT Press, 1992.

Jerry Fodor, *The Mind Doesn't Work That Way: The Scope and Limits of Computational Psychology*, The MIT Press, 2001.

Jaegwon Kim, *Philosophy of Mind*, Westview Press, Third Edition, 2010.

Jaegwon Kim, *Mind in a Physical World: An Essay on the Mind-Body Problem and Mental Causation (Representation and Mind)*, The MIT Press, 2000.

Graham Macdonald, David Papineau (eds.), *Teleosemantics*, Oxford University Press, 2006.

Ruth Millikan, *Language, Thought, and Other Biological Categories: New Foundations for Realism*, The MIT Press, 1987.

Ruth Millikan, *White Queen Psychology and Other Essays for Alice*, The MIT Press, 1995.

David Papineau, *Thinking about Consciousness*, Oxford University Press, 2004.