

Philosophy of Science
Философия на науката

МЯСТОТО НА ИНТЕЛЕКТА В ТЕОРИИТЕ ЗА СЪЗНАНИЕТО ОТ ПО-ВИСОК РЕД (ТСПР)

Давид Пеев

Софийски университет „Св. Климент Охридски“

Резюме. Настоящата статия има за цел да внесе яснота по отношение на мястото, което терминът „интелект“ заема в концепциите за ум и съзнание, и по точно – в теориите за съзнанието от по-висок ред. В резултат се надявам да достигна до добре обоснован извод, засягащ връзката между интелекта, от една страна, и еволюцията и адаптацията на организмите, от друга. Макар че тази заключителна формулировка може да се разглежда по-скоро като работна дефиниция на интелекта, последната би могла да фасилитира по-нататъшни изследвания по темата.

Ключови думи: интелект; ум; теории за съзнанието от по-висок ред; еволюция; адаптация

Философските дебати около интелекта през последните години са по-скоро фокусирани около концепцията за изкуствен интелект. Това допускане може да се отнесе както към опити да се анализира понятието за интелект, като се сравни със или различи от изкуствено създадена информационна система (виж напр. Searle, 1980; Dretske, 1993; Leg & Hutter, 2007 и др.), така и към многобройните етически дискусии, които възможността за изкуствен интелект създава (Crombag, 1993; Moor, 1985, 2006; Pereira & Saptawijaya, 2016; Ashrafiyan, 2015 и др.). Проблемът е, че понятието „интелект“ към момента не може да се обяви за научен термин. Могат да се открият множество различни дефиниции на интелигентност, а това води до липсата на ясен референт, по отношение на който да се започне една продуктивна дискусия. По тази причина настоящият текст почти не се занимава с понятието за изкуствен интелект, а се връща една стъпка назад – към въпроса „Какво изобщо следва да се разбира под „интелект“ и „интелигентност““. Нуждата от връщане към този по-първичен и изначален проблем е засегната също от Лег и Хътър, според които ние не просто не разполагаме с единна концепция за интелигентност, а опитите ни да създадем такава, ни водят до редица други понятия, които също не сме дефинирали успешно (Leg & Hutter, 2007: 392).

Проблемът за дефиницията на „интелигентност“ е обвързан пряко и с изясняването на носителите на последната. Кои същества попадат в класа от интелигентни системи и на базата на какъв критерий? Защо говорим за интелигентни хора, може би дори за интелигентни кучета, но никога не описваме камъните като интелигентни? Фактът, че интелигентността стандартно се приписва на живи същества, неизбежно води до проблема тяло – ум (mind-body) и нуждата от прецизиране на референцията на „интелект“ с оглед на ума или тялото. Също като „интелект“, думите „ум“ и „съзнание“ страдат от един съществен терминологичен недостатък – тяхната многозначност. Липсата на консенсус около това към какво реферираме, когато говорим за съзнание, дава основание да се твърди, че понятието за съзнание е понятие мелез (срв. Block, 2002: 206). Ето защо една легитимна изследователска цел е да се открие съвременна теория за съзнанието, която да бъде съвместима с множеството приложения на термина „интелект“. С оглед на философията на науката, такова позициониране на „интелект“ би имало значимост, доколкото предлага базова методология за бъдещи изследвания в областта на еволюцията на интелекта. Изискването към теорията за съзнанието, която се търси, е да е съвместима с еволюционната теория за произхода на видовете, доколкото последната е единственият смислен вариант за отговор на въпроса за природата на *homo sapiens* и други, „нечовешки“, животни, притежаващи съзнание. Ще защитавам умерен редукционизъм по отношение на ума/съзнанието, като стремежът ми е да покажа, че именно натуралисткото разбиране на интелекта е най-съвместимо с методите и резултатите на съвременната наука.

Постигането на терминологично разбиране за „интелект“ изисква да се посочи дали интелектът следва да се разглежда понятийно, или предикативно. Нужно е кратко разяснение: „понятие“ използвам най-вече като термин на менталистките теории за понятията, като например хипотезата за езика на мисълта (срв. Harman, 1973; Fodor, 1975). Изключвам редукционисткото разбиране за понятията, при което се отхвърля тезата, че понятията са ментални партикуларии и се идентифицират с определени способности на индивида (Millikan, 2000; Brandom, 1994). В настоящата таксономия разглеждането на понятията като способности се причислява по-скоро към предикативното разглеждане на интелекта. Така, ако за интелект се говори като за понятие, това би имало определени импликации за начина, по който интелектът се представя в рамките на една по-широка теоретична рамка. Пример за такава импликация е, че единствено разгледан понятийно, терминът „интелект“ може да реферира към психологическо свойство на даден индивид. От друга страна, предикативното разглеждане позволява предикатът „е интелект“ да се приписва на уменията на дадена система да обработва информация (където не се изисква системата да притежава каквато и да било Аз-овост и психологически свойства). Предикативното разбиране е по-съвместимо с опита да се

обособи терминът „интелект“, защото позволява той да се приложи в по-голям брой от случаи. Също така, предикативното разбиране кохерира с предпоставката за еволюционно изменение на интелекта, доколкото не изисква той да се обвързва със специфично свойство на системата. Поради тези две основания в настоящия текст интелектът ще се разглежда като предикат, а не като понятие.

Прокарването на разграничение между различните типове организми в съвременните теории за ума следва да отговаря на доста по-строги критерии и да се съобразява с емпиричните данни, предоставени от биологията, неврофизиологията и невронауката. Така идеята на Кант за разделянето на перцептивните способности на една биологична система от нейната по-висша, мисловна дейност се развива драстично, за да достигне еманацията си в т.нар. теории за съзнанието от по-висок ред (ТСПР). Създаването на тези теории е мотивирано от различни казуси, свързани с начина, по който светът се възприема от по-нисшите и по-висшите организми. За да демонстрирам какво стои в основата на една ТСПР, ще реконструирам пример на Дрецке, като добавя леки изменения, които имат за цел по-добро онагледяване на разликите между различните свойства на съзнанието. Да допуснем, че се намирам в центъра на една стая заедно с друг къщен обитател – мишката Джери. В ъгъла на стаята млада жена свири на пиано. Разбира се, и аз, и Джери преживяваме (сетивно долавяме) звука на пианото. Аз обаче мога: а) да преживявам пианото като „пиано“; б) да формулирам пропозицията „Тя свири на пианото“ и да мисля за пианото, а това се дължи на факта, че разбирам концепцията за пиано. Джери не притежава такава концепция и едва ли може да формира каквито и да било убеждения по отношение на пианото, подобни на моите (Dretske, 1997). Разликата между мен и Джери е, на първо място, в: а) това, че аз притежавам феноменално съзнание за пианото. ТСПР са построени върху идеята, че моето феноменално съзнание за пианото изисква някаква „осведоменост“ (awareness) за състоянието, в което съм, която е от по-висок ред и липсва у една мишка (Rosenthal, 2004: 17). По интересен обаче е въпросът дали умението ми да правя б) не стои в основата на а). От една страна, може да се допусне, че аз не бих могъл да бъда във феноменално съзнателно ментално състояние, освен ако на него не съответства предварително някакво състояние от по-висок ред (да речем – притежаване на концепция за това какво е пиано) (Lucan, 2001). От друга страна, може би именно животните ни дават примери за феноменално съзнание без наличието на ментални състояния от по-висок ред (срв. напр. Туе, 1995: 5 – 6).

За настоящата статия особен интерес представлява да се определи как интелектът, като способност, се включва в разграничаването на различни свойства на съзнанието или дори различни типове съзнание. Сред ТСПР теоретиките се водят редица дебати, но аз ще се фокусирам само върху два конкретни

проблема, които според мен имат пряко отношение към локализирането и определянето на функциите, които изпълнява интелектът в една биологична система. Първо, това е въпросът дали даден тип съзнателни процеси може да съществува отделно от друг. Ако отговорът е да, следва, че може да се говори за „чисти“ феноменално съзнателни и „чисти“ достъпно съзнателни (access conscious) състояния (Block, 2002). Второ, дали тези състояния от по-висок ред, които са състояния за възприятията на системата, следва да бъдат перцептивни, или рефлексивни? Първия тип теории ще наричам Теории за Перцепциите от По-висок Ред (ТППР), а втория – Теории за Мислите от По-висок Ред (ТМПР). В ТППР дадено състояние е съзнателно (е състояние за), ако е съпроводено от *вътрешно усещане* (inner sense) от по-висок ред (Armstrong, 1968). Понятието на Армстронг за вътрешно усещане е заимствано от емпиризма на Лок, но се обяснява в репрезентационистки термини. В случая с ТМПР, за да бъде една перцепция съзнателна, индивидът трябва да съзнава, че притежава такава съзнателна перцепция (Rosenthal, 1993). Така „мислите“ играят ролята на условие за осъзнаване на възприятията. Тук „мисъл“ се използва *sensu latissimo*, като, най-грубо казано, терминът обозначава убежденията, знанията на индивида или друг тип концептуални репрезентации (Dretske, 1997: 106).

Обвързването на интелекта с еволюционната теория прави първия от гореспоменатите проблеми особено важен. Когато става дума за съзнанието на дадено същество¹⁾, то следва да е възникнало в резултат на естествен отбор. Самата селекция предполага, че съзнанието е продукт на адаптация и неговите проявления следва да имат роля в контрола на поведението на дадения организъм (Cairns-Smith, 1996: 279). Да допуснем, че феноменалното съзнание и съзнанието за достъпа са две различни проявления на съзнанието, отговарящи на два различни процеса. Тогава е възможно да се спекулира, че съзнанието за достъпа е продукт на по-високо ниво на адаптивност (да предлага по-ефективен контрол върху поведението). Интелектът трябва или да бъде отнесен към това „съзнание за достъпа“, или ако бъде отнесен към всички съзнателни процеси, да е оправдано говоренето за степени на интелигентност. Така, ако се ползва метафората на Дрекке, в първия случай интелектът ще е като „богатство“. Той ще може да бъде приписан само на малък брой сложно устроени организми. Ако разбирането ни за интелект обаче е „по-широко“, той ще бъде като „парите“ – всички ги имат в някаква степен, просто някои притежават повече (Dretske, 1993: 201). За да работи тази метафора изобщо обаче, трябва да има разлика между системи, които имат съзнание за това „какво е да си“, и системи, които могат да правят заключения и *рационално* да контролират поведението си (Block, 1990)²⁾.

Обичайно се счита, че примерите на Блок за случаи на феноменално съзнание без съзнание за достъп са достатъчно убедителни, но дали това е така. Един от примерите, посочени от него, е сънуването, което може изцяло да се опише като съзнаване за това какво човек вижда/чува/преживява, без да е в

състояние да контролира тези преживявания или да прави логически изводи, базирани върху тях (Block, 2002: 213). Емпиричните свидетелства обаче не подкрепят този пример. Още през 1995 г., Ревонсуо се опитва да демонстрира, че емпиричните данни, свързани с поведението на агенти в сънно състояние, посочват, че то не може да бъде описано като съзнателно (Revonsuo, 1995). Само две години по-късно неврофизиологът Джоузеф Боген се изказва в подкрепа на тезата, че при процеса на сънуване липсват както сензорни входящи сигнали, така и поведенчески изходни реакции на организма (Bogen, 1997: 144). Дори този пример да е опроверган, Блок настоява, че феноменалните свойства на съзнанието не могат да бъдат идентифицирани с функционални свойства на нервната система³⁾. Така се стига до друг пример за „феноменално съзнание“, а именно – „да знам какво е един квадрат да е червен“ (Block, 2002: 206). Въпреки настояването си, че „ще стои далеч“ от случаи като стаята на Мери (Jackson, 1986), които според него са проблематични, тук Блок попада в същия капан с оглед на резултатите, предоставени от съвременната наука. Целта му е да покаже, че феноменалното свойство на съзнанието създава „обяснителната пропаст“ между процесите на нервната система и свойствата на съзнанието още по-голяма. Проблемът е, че въвеждането на този пример обосновава твърдението, че тази пропаст е изкуствено „разширена“ от самия Блок. Има разлика между твърдението, че „има нещо, което е да преживяваш червен квадрат“, и това, че „има нещо, което е да си функционално работещ мозък“. Разликата се състои в това, че първото изисква да се приеме предпоставката, че има нещо, което е червено (Van der Heijden, Hudson and Kurvink, 1997: 157). Също като при стаята на Мери, примерът на Блок, изглежда, допуска грешката да субстанциализира цветовете, без да приведе каквито и да било свидетелства за този аргументативен ход. Това води до приемане на положението, че преживяването на цветовете е ясен пример за феноменално съзнание на базата на *ad hoc* грешка в аргумента. Такъв тип грешка не е чужда и на материалистите, както отбелязва Хардин:

„Така онези, които настояват, че цветовете имат феноменална природа, не правят почти нищо, за да обяснят феноменологията им, а „научните“ материалисти не обръщат много внимание на науката. Дебатът продължава благодарение на концептуален анализ и посредствени аргументи, но базата данни е почти същата, която е била налична още по времето на Джон Лок. Резултатът е, че проблемът за онтологичния статус на цветовете не просто намира същите решения, които и в XVIII век, но въпроси, свързани със сходството и взаимното изключване между цветовете, са в задънена улица. Отвъд това, общото разбиране за това как преживяването за цветовете се свързва с наименованията за тях, е забулено в обкръване“ (Hardin, 1988: xxi).

Приемам патоса на Хардин и смятам, че е редно проблемът за интелекта да се разреши на базата на интердисциплинарно взаимодействие между филосо-

фията и емпиричните науки. Първо обаче следва да се върна към въпроса за „чистите“ феноменологични състояния на съзнанието. Внимателното изследване на феномени като „излюзията с водопада“ показва, че едни и същи стимули могат да бъдат обработени от нервната система по множество начини, като дори когато група от неврони си „почива“, тя активно пресъздава картината за течащия водопад (Hardin, 1988: 110). Следователно така наречените „рационални процеси на контрол“, идентифицирани със съзнанието на достъпа, протичат дори при чисто визуалното преживяване на цветове, защото нервната система селектира и контролира най-оптималния начин да предостави информация от получените стимули. Обособяването на феноменални състояния на съзнанието, в които не протича подобен контрол, изглежда, е наивно.

Сега към втория въпрос – къде следва да бъде позициониран интелектът с оглед на разграничението между феноменално съзнание и съзнание за достъпа? Въз основа на изложения по-горе аргумент изглежда, че самото разграничение би следвало да отпадне. Нещо повече, анализът на еволюцията на интелекта не се нуждае от обвързване на интелекта с каквото и да било преживяване на kvalii. Хроматичните перцептивни състояния например могат успешно да бъдат идентифицирани със състояния на нервната система, което прави представата за цвят илюзорна, но полезна (ibid, 111). Възприемането на цветове е адаптация, която също като други подобни илюзии в човешкия свят води до по-ефективно поведение. Несъзнателното превръща не на множество сигнали от външната среда в единна съзнателна представа за цвят прави човешкия организъм по-способен на ориентация в определени ситуации и по-ефективен в уменията му да разграничава конкретни елементи от заобикалящата го среда. Интелигентното поведение съответства както на съзнателни мозъчни процеси, така и на несъзнателните, на които обикновено бива противопоставяно. Така примерите за „феноменални състояния на съзнанието“ могат да водят до интелигентно поведение, доколкото повишават адаптивността на биологичната система, в която възникват.

Дори и да се възрази, че примери за обяснение на цветовете, заимствани от физиката, не са релевантни за дебата около природата на съзнанието, други примери, предоставени от съвременната невронаука, със сигурност са. Те показват, че цветовете не са единствената „илюзия“ на съзнанието ни. Могат да се търсят корелации между множество ментални състояния със състояния на нервната система. Установено е например, че прилагането на електрически сигнали към мозъка е в състояние да произведе вариации в съзнанието, като образи и/или спомени (вж. напр. Penfield & Perot, 1963: 595). Обосновимо е да се допусне и обратното: електрически сигнали биват регистрирани в мозъка за всяко съзнателно усещане. Нещо повече, експериментите на Бенджамин Либет ясно демонстрират как електроенцефалограмите на тествания субект регистрират активност преди съзнанието за усещането. С други думи, не е коректно да

се твърди, че съзнание за достъпа „автономно“ участва в процеса на вземане на решения (вж. напр. Igalman, 2017). Това не просто не отговаря на емпиричните данни, но те сочат, че има „прозорец“ от (поне) 60 – 70 милисекунди, преди мозъкът да е в състояние да изчисли дори най-простото усещане (вж. Libet, 1985: 529 – 566). Този времеви интервал е твърде голям, за да бъде интерпретиран като „феноменално преживяване“. Струва ми се невероятно да е оправдан такъв голям интервал просто за да може съзнанието да „избере“ дадено състояние в процеса на невротрансмисия. Когато се приписва интелект на дадена система, това се прави въз основа на целокупната ѝ организация, на всички процеси и подпроцеси, които системата изпълнява. Така архаичната представа за „интелект“, „разсъдък“, „ум“, представени като нещо отделно от възприятията, следва да бъде отхвърлена. Ако „рационалните човешки същества“ превъзхождат останалите биологични видове, това се дължи по-скоро на тяхната по-сложно устроена в резултат на еволюцията нервна система.

БЕЛЕЖКИ

1. Има разлика между това да се твърди, че дадено същество действа съзнателно (creature consciousness), и това дали определено ментално състояние, в което се намира съществото, е съзнателно (state consciousness) (Manson, 2000).
2. В случая „рационалният“ контрол над поведението се съпоставя с рефлексивни действия, извършвани от по-нисши организми. Сензорно възприемане на стимули като топлината у бълхите например едва ли може да се обясни като целенасочено (рационално) (срв. McFarland, 1981: 449).
3. Тази версия на дуализма е по-известна като „дуализъм на свойствата“ и както беше отбелязано, почива на допускането, че всяко преживяване има своя субективна природа (квалия).

REFERENCES

- Igalman, D. (2017). *Mozakat. Tova si ti*. Sofia: Ciela.
- Armstrong, D., (1968). *A materialist theory of the mind*. London: Routledge.
- Ashrafian, H. (2015). Artificial Intelligence and robot responsibilities: Innovating beyond rights. *Science and Engineering Ethics*, 21 (2), pp.317 – 326.
- Block, N. (1990). Consciousness and accessibility. *Behavioral and Brain Sciences*, 13, pp. 596 – 598.
- Block, N. (2002). Concepts of consciousness. In David J. Chalmers (ed.), *Philosophy of Mind: Contemporary Readings*. Oxford University Press. pp. 206 – 218.
- Bogen, J. (1997). An example of access-consciousness without phenomenal consciousness?. *Behavioral and Brain Sciences*, 20: 1, p. 144.

- Brandom, R. (1994). *Making It Explicit: Reasoning, Representing, and Discursive Commitment*. Cambridge, MA: Harvard University Press.
- Cairns-Smith, A. G. (1996). *Evolving the mind: On the nature of matter and the origin of consciousness*. Cambridge, England: Cambridge University Press.
- Crombag, H. (1993). On the artificiality of artificial intelligence. *Artificial Intelligence and Law*, 2 (1), pp. 39 – 49.
- Dretske, F. (1993). Can intelligence be artificial? *Philosophical Studies*, 71 (2), pp. 201 – 16.
- Dretske, F. (1995). *Naturalizing the mind*. Cambridge, MA: MIT Press.
- Fodor, J. (1975). *The Language of Thought*. New York (NY): Thomas Y. Crowell.
- Hardin, C. (1988). *Color for Philosophers: Unweaving the Rainbow*. Hackett Publishing Company, Inc.
- Harman, G. (1973). *Thought*. Princeton, NJ: Princeton University Press.
- Jackson, F. (1986). What Mary Didn't Know. *The Journal of Philosophy*, 83 (5), pp. 291 – 295.
- Leg, S. & Hutter, M. (2007). Universal intelligence: A definition of machine intelligence, *Minds and Machines*, 17 (4), pp. 391 – 444.
- Libet, B. (1985). Unconscious cerebral initiative and the role of conscious will in voluntary action. *The behavioral and brain sciences*, 8, pp. 529 – 566.
- Lycan, W. (2001). A simple argument for a higher-order representation theory of consciousness, *Analysis*, 61: 1, pp. 3 – 4.
- Manson, N. (2000). State consciousness and creature consciousness: A real distinction. *Philosophical Psychology*, 13 (3), pp. 405 – 410.
- McFarland (1981) *The Oxford companion to animal behavior*. Oxford: Oxford University Press.
- Millikan, R. (2000) *On Clear and Confused Ideas*. Cambridge: Cambridge University Press.
- Moor, J. (1985). What is computer ethics?. *Metaphilosophy*, 16.4, pp. 266 – 274.
- Moor, J. (2006). What is computer ethics?. *IEEE Intelligent Systems*, 21.4, pp. 18 – 21.
- Penfield, W. & Perot, P. (1963). The brain's record of auditory and visual experience: a final summary and discussion. *Brain*, 86 (4), pp. 595 – 696.
- Pereira, L. M. & Saptawijaya, A. (2016). *Programing machine ethics*. Switzerland: Springer International Publishing.
- Revonsuo, A. (1995). Consciousness, dreams and virtual realities. *Philosophical Psychology*, 8 (1), pp. 35 – 58.
- Rosenthal, D. (1993). Higher-order thoughts and the appendage theory of consciousness. *Philosophical Psychology*, 6 (2), pp. 155 – 66.

- Rosenthal, D. (2004). Varieties of higher-order theory. In: Gennaro, R. (ed) *Higher-Order Theories of Consciousness: An Anthology*. Amsterdam: Johns Benjamin Publishing Company, pp. 17 – 44.
- Searle, J. (1980). Minds, brains, and programs. *Behavioral and Brain Sciences*, 3 (3), pp. 417 – 57.
- Tye, M. (1995). *Ten Problems of Consciousness: A Representational Theory of the Phenomenal Mind*. Cambridge, MA: MIT Press.
- van der Heijden, A. & Hudson, P. & Kurvink, A. (1997). On widening the explanatory gap. *Behavioral and Brain Sciences*, 20:1, pp. 157 – 158.
- Wimsatt, W. (2006). Reductionism and its heuristics: Making methodological reductionism honest. *Synthese*, 151 (3), pp. 445 – 475.

“THE INTELLECT” ACCORDING TO THE HIGHER-ORDER THEORIES OF CONSCIOUSNESS

Abstract. The central aim of this paper is to provide certain clarity regarding the question ‘What role does ‘the intellect’ play from the perspective of a Higher-order Theory (HOT) of consciousness. My hope is to show that there are grounds for the conclusion that there is a close relation between what we call ‘intellect’ and the evolutionary and adaptive processes that a biological system undergoes. As for the final result, my goal is admittedly modest – to merely outline a working definition of the term ‘intellect’, which is, however, compatible with evolutionary theory. Nevertheless, I hope that, if successful, such definition can facilitate future research in the field.

Keywords: intellect; mind; higher order theories of mind; evolution; adaptation

✉ **Mr. David Peev, Assist. Prof.**

Faculty of Philosophy

University of Sofia

E-mail: david_peev@phls.uni-sofia.bg

Philosophy Teacher in the American College in Sofia

E-mail: d.peev@acsb.org