Роджър Пенроуз ♦ Емануеле Северино
Фабио Скардили ♦ Инес Тестони ♦ Джузепе Витиело
Джакомо Мауро Д'Ариано ♦ Федерико Фаджин

ИЗКУСТВЕН СРЕЩУ ЕСТЕСТВЕН ИНТЕЛЕКТ

София, 2024

First published in English under the title
Artificial Intelligence Versus Natural Intelligence
by Roger Penrose, Emanuele Severino, Fabio Scardigli, Ines Testoni, Giuseppe Vitiello, Giacomo Mauro D'Ariano, Federico Faggin and Fabio Scardigli, edition: 1
(Editor Fabio Scardigli)

ISBN 978-619-01-1482-6

Роджър Пенроуз
Емануеле Северино
Фабио Скардили
Инес Тестони
Джузепе Витиело
Джакомо Мауро Д'Ариано
Федерико Фаджин

ИЗКУСТВЕН СРЕЩУ ЕСТЕСТВЕН ИНТЕЛЕКТ

Съставител
Фабио Скардили

Превод от английски
Росен Люцканов

ИЗТОК ЗАПАД

Съдържание

Увод

Фабио Скардили*

Тази книга съдържа транскрипция на изказвания и обсъждания между Роджър Пенроуз и Емануеле Северино по време на конференцията „Изкуствен срещу естествен интелект“, проведена в Милано, в конгресния център „Карипло“, на 12 май 2018 г.

В добавка към пленарните доклади на Пенроуз и Северино са изнесени лекции от Джузепе Витиело (специалист по теоретична физика), Мауро Д'Ариано (специалист по теоретична физика) и Инес Тестони (психолог), които са основа на текстовете, поместени в тази книга.

Конференцията е замислена и организирана (подобно на предишната – „Детерминизъм и свободна воля“) от група приятели и колеги: Фабио Скардили, Марсело Еспозито и Марко Доти. Изказваме благодарност на нашия колега Масимо Блазоне, който ни помагаше по време на конференцията.

Успехът на конференцията беше поразителен, дори по-голям от този на предишната. Над 600 души се струпаха в главната зала на конгресния център „Карипло“ и в двете съседни поме-

* Математически факултет, Миланска политехника, Милано, Италия, e-mail: fabio@phys.ntu.edu.tw. (Всички бележки под линия, без изрично упоменатите, са на авторите.)

щения, оборудвани с видеоекрани. Това нагледно показва големия интерес на обществото към темите за изкуствения интелект, теориите за съзнанието, интелигентните устройства и т.н.

1. Разбиране и алгоритми

Що се отнася до въпросите, обикновено причислявани към темата за изкуствения интелект, гледните точки на двамата основни докладчици, специалиста по математическа физика Роджър Пенроуз и философа Марсело Северино, очевидно са доста различни. Въпреки това читателят скоро ще установи, че и двамата са на мнение, че все още не разполагаме с „интелигентни устройства“ и че ако се придържаме към гледната точка на т.нар. силен изкуствен интелект (понастоящем широко приета), никога няма да можем да създадем такива устройства. Това мнение се подкрепя също от авторите на други текстове в книгата, макар и по специфичен за тях начин.

В своя доклад Пенроуз се фокусира върху понятията „интелект“, „разбиране“ и „съзнание“. Тъй като е математик, той се шегува, че „връзките между думите са по-важни за мен от тяхното „истинско“ значение“. Иначе казано, връзките между понятията са по-показателни от съдържателните им дефиниции. Ето защо Пенроуз изхожда от идеята, че думата „интелект“, поне в стандартната ѝ употреба, е свързана с „разбиране“, а „разбиране“ изисква „осъзнаване“ или „съзнание“. Преминавайки през редица примери, които обсъжда подробно, Пенроуз показва, че машините или компютърните програми, с които разполагаме днес, са изчислителни устройства, които, макар да са усъвършенствани, по същество се основават на идеалния първообраз на машината на Тюринг. От своя страна „интелект“ и „разбиране“, изглежда, се характеризират чрез свойства, които не се свеждат до изчислимост. Интелектът е нещо повече от способност за извършване на изчисления. Примери, свързани с шаха, математическата индукция (вълнуващата теорема на Гудстийн), паркетирането на евклидовата равнина (чрез полиомино, което не може да бъде

постигнато чрез изчислителни алгоритми), демонстрират, че „разбирането е нещо, което не се постига чрез правила". Следователно общо качество на разбирането е това, че то не се свежда до алгоритъм, и според Пенроуз то не е нито част, нито продукт на (изключително) сложна система от правила (алгоритъм).

2. Квантова механика и съзнание

След това Пенроуз предлага обзор на двата теоретични фундамента на съвременната физика, а именно общата теория на относителността и квантовата механика, след което посочва, че има само един компонент на днешната физика, който не може да бъде възпроизведен чрез компютър, тъй като не е изчислим. Това е процесът на измерване според квантовата механика, поточно така нареченият „колапс на вълновата функция". Той не се описва чрез уравнението на Шрьодингер и не може да бъде имплементиран (дори принципно) чрез изчислителен алгоритъм.

По думите на Пенроуз: „Идеята е, че колапсът е процес, който не е изчислим. В действителност той по нещо напомня за свободната воля, тъй като според днешната физика сам определя курса си. Някак решава дали да е тук, или там."

Сред всички физични теории и физични явления, изглежда, има само две неща, които в същността си са неалгоритмични и неизчислими: процесът на измерване в квантовата механика (колапсът на вълновата функция) и феноменът на „разбирането" или „осъзнаването", с което се отличава (човешкото) съзнание. В работите си Пенроуз отдавна се опитва да свърже тези две понятия. За пръв път в книгата си „Новият разум на царя"* той формулира теория, според която началото на неалгоритмичните процеси на „разбирането", „осъзнаването" и самото „съзнание" трябва да бъде търсено в определени квантови процеси, протичащи в конкретни области в мозъка. Благодарение на сътрудни-

* Пенроуз, Р. *Новият разум на царя*. София, университетско издателство „Св. Климент Охридски", 1998. – Б.пр.

чеството му с анестезиолога Стюарт Хамероф през 90-те години на XX век е идентифициран обещаващ кандидат за място, където протичат „спонтанни оркестрирани редукции", така наречените „микротубули". Микротубулите са изключително малки субклетъчни структури, вложени в невроните – в аксоните и дендритите.

В тези структури се поддържа достатъчно дълго квантово кохерентно състояние, което позволява на квантовия колапс на вълновата функция да влияе пряко на „елементите" или „атомите" на квантовото съзнание, които можем да определим като „протосъзнание". Според Хамероф и Пенроуз това може би са градивните единици, от които е изградено съзнанието. Очевидно е, че тези градивни единици нямат осъзнати цели, нито планове, а още по-малко разбират значения, но от тях могат да възникнат структури, способни на съзнателно поведение.

3. Протосъзнание

Пряко следствие на тази теория е, че съзнанието далеч не е само човешка характеристика и възниква винаги, когато са налице структури като микротубулите, съответно се среща при животни като човекоподобните маймуни, делфините, кучетата, котките или мишките. Това заключение има очевидни етически импликации и тъкмо то провокира дискусията със Северино. Освен това Пенроуз накратко обсъжда възможността за „конструиране" на съзнателни, съответно интелигентни устройства. От негова гледна точка това може да бъде постигнато единствено чрез агрегиране на елементи, надарени с протосъзнание, и осигуряване на подходяща среда, в която да протече „оркестрирана обективна редукция" на вълновата функция.

Според Пенроуз, доколкото се основават на чисто алгоритмични процеси, днешните компютри, а вероятно и утрешните, очевидно са и ще си останат лишени от съзнание, съответно са неспособни да разбират истински и да проявяват интелигентност. Няма опасност един ден „интелигентни машини" да пре-

вземат света, заплашвайки с унищожение човешката раса. Известно е, че по този въпрос Стивън Хокинг имаше тъкмо обратното мнение.

4. Свободна воля и сингулярности

Във връзка с въпроса за свободната воля и в светлината на казаното можем да направим още някои изводи. Както е известно, специалистите по обща теория на относителността се ужасяват от сингулярностите, които се появяват в тази теория. Роджър Пенроуз спечели Нобеловата награда за физика през 2020 г. тъкмо заради основополагащата си статия от 1965 г., в която е доказана теорема, според която при определени условия (без да е необходима сферична симетрия), според общата теория на относителността, сингулярностите са често срещано и неизбежно явление (както в миналото – космологичната сингулярност на Големия взрив, така и в бъдещето – сингулярностите на черните дупки).

Защо обаче специалистите по теоретична физика не обичат сингулярностите?

Стандартният отговор гласи, че когато се появи сингулярност, способността на теорията да предсказва бъдещето изчезва напълно. От сингулярността може да се появи буквално всичко, при това по напълно непредсказуем начин, тъй като физичните закони по дефиниция са неприложими към нея. За да защити наблюдаемата вселена от такива чудовища, Пенроуз формулира преди време хипотезата за космичната цензура, според която сингулярностите винаги остават скрити зад хоризонт на събитията и съответно не могат да влияят на вселената извън тях.

Според разсъжденията на Пенроуз има и друг феномен, който ни сблъсква със „сингулярност“, при която теорията се оказва напълно лишена от способността да предсказва: ключовото събитие в квантовата механика! Всеки колапс на вълновата функция води до крах на изчислимостта, съответно до „сингулярности“, при които квантовата теория се оказва неспособна

да ни осигурява предвиждания. Например, ако вземем излъчването на отделен фотон от атом във възбудено състояние, няма да е възможно да предскажем нито кога, нито в каква посока ще се осъществи то. Нещо аналогично се случва при разпада на неутрон: за нито едно такова събитие не знаем нито кога, нито в каква посока ще бъде излъчена двойка от неутрино и електрон. Квантовата теория ни осигурява единствено вероятностни разпределения, които, макар и да се съгласуват напълно с експерименталните данни, по дефиниция се отнасят до класове от събития, а не до отделни събития. Съответно „сингулярностите", случаите, в които предсказуемостта изчезва, според квантовата механика са навсякъде около нас.

5. Конструиране на съзнание

Две от темите, към които Емануеле Северино насочва своята критика, засягат „мястото", където е локализирана човешката интелигентност или съзнанието, и възможността за „конструиране" на интелигентно (или съзнателно) устройство.

В началото Северино се позовава на хипотетико-дедуктивния характер на науката. Всички науки – и в частност математизираните – се основават на постулати, иначе казано, на твърдения, които се приемат без доказателство и от които следват според определени (логически) правила други твърдения. Самите постулати и дори дедуктивните правила не се смятат за неопровержими истини дори в рамките на самата наука: те не са нищо повече от „конвенции". Или ако използваме термин, на който Северино приписва по-широко значение, те са „положения на вярата". Иначе казано, те са израз на „волята за власт". Според Северино „изборът между две конкуриращи се теории в крайна сметка се определя от това коя от тях е способна да преобрази света". Самата наука признава, че не е възможно да достигнем знание, което никой, нито боговете, нито хората, не е способен да отрече. Затова според Северино науката не цели да достигне неопровержими истини, а по-скоро да наложи контрол над света.

Тази цел според Северино е противоположна на целта, която поначало е основната цел на философията – причината тя да се роди преди 2500 години в Древна Гърция, – това да бъдат разбулени неопровержимите истини.

6. Математическо моделиране

От друга страна, тезата на Северино, че древните гръцки математици и последователите им чак до Галилей се опитват да достигнат до неопровержими, епистемично гарантирани истини („да познават теоремите, както ги знае Бог", по думите на Галилей), със сигурност подлежи на критика. Всъщност е съвсем ясно, особено според най-новите исторически реконструкции (вж. напр. Лучо Русо), че понятието за „математически модел" е било налично още в работите на Евклид, Архимед и други математици от същия период. Свободният избор на постулатите, възможността да си „играеш" с тях, за да откриеш по-добър модел на дадения феномен, или да вариращ даден набор от (математически) твърдения са операции, които са ясно формулирани и ефективно осъществявани от древните математици, но и от бащите на научната революция в епохата на модерността Коперник, Галилей, Нютон и т.н. Може да се твърди, че нито един от тях в действителност не е вярвал в постигането на абсолютната истина, а вместо това те са се занимавали с (математически) модели, които са по-добри от възприетите по-рано (например с търсенето на физика, която е по-добра от тази на Аристотел, или на астрономия, която превъзхожда тази на Птолемей). (В тази връзка виж увода на сборника от конференцията „Детерминизъм и свободна воля".)

7. Явяване на света

След това Северино въвежда понятието за „явяване на света" – онова измерение, в което се случва всичко, в което всичко намира изява, измерението, от което черпим сведения за всеки

отделен факт или нещо: „Няма и една стъпка, която науката е в състояние да направи, която не извира от явяването на света." Според Северино „явяването на света" не е нещо наред с всичко останало, тъй като съдържа всички минали, настоящи и бъдещи неща в света, всичко, до което имаме достъп.

Северино се позовава на идеята за „явяване на света" и във втората част от лекцията си, в която се противопоставя на възгледа на Пенроуз относно „мястото [в мозъка], където да търсим съзнанието". Според Северино търсенето на място, „където се помещава съзнанието", всъщност предполага третирането на съзнанието като конкретен аспект на „явяването на света", което следва да се разглежда като първичната форма на съзнанието.

Въпреки това идеята за „съзнанието на света", изглежда, споделя някои белези с „атомите" на протосъзнанието (споделяни от различни обекти), която е въведена от самия Пенроуз. Сходства могат да бъдат открити и с други понятия, обсъждани от автори в книгата – Витиело, Д'Ариано и Фаджин, макар и с различен акцент и интерпретирани от друга гледна точка.

8. Производство

В ядрото на критиката на Северино, отправена срещу идеята за „изкуствен" интелект, или „изкуствено" съзнание, е идеята за „производство". Както е типично за философската позиция на Северино, грешката, основният нихилистичен пропуск на западната философия (и цивилизация), се крие в глагола „произвеждам". Макар и на пръв поглед съвсем невинна, идеята за производство, *poiesis* на гръцки, крие в себе си схващането, че нещо може да бъде създадено от нищо и (при желание) да бъде сведено обратно до нищо. Според Северино това, че нещата могат да бъдат създавани и унищожавани, убеждението, че те блуждаят между съществуване и несъществуване, съответно „ставането" на нещата, което за нас е съвършено очевидно, всъщност не се потвърждава от наблюдението, а е част от теория, която е само една от възможните интерпретации на ре-

алността наред с много други. Зад това упорито убеждение се крие призракът на нихилизма: убеждението, че всяко едно нещо всъщност е нищо.

Ясно е, че според тази гледна точка идеята да се произведе изкуствен интелект не е нищо повече от проява на дълбокия нихилизъм, който е обзел западната цивилизация.

9. Изкуствен интелект

Северино не спира дотук. Той констатира, че разбирането на Платон за „производството" като „причина, която кара нещо да премине от несъществуване в съществуване", прониква в западния начин на мислене като цяло – във философията, икономиката, правото и математическите науки, и неизменно се третира като самоочевидна. Освен това западният начин на мислене третира Човека като единственото същество, способно да организира нужните средства за целите на производството. Както добавя Северино в тази връзка, точно това е дефиницията за машина! Единствената разлика е, че за момента машините не притежават цели: „Човекът е машина, която организира средствата с оглед на производството на цели, имайки предвид тъкмо наличието на целите, идеалното им присъствие." По този начин Северино достига провокативния извод, че „природният" Човек се мисли като машина, а самият свят – като механизъм, чрез който средствата се организират с оглед на производство на цели. Съответно, имайки предвид начина, по който се разбира Човекът на Запад, Човекът, или по-скоро неговата природна интелигентност, „е" поначало „изкуствена" интелигентност.

Противопоставяйки се на тази теза, Северино твърди, че „явяването на света като цяло – тази първична и фундаментална форма на съзнанието – не може да бъде произведено, ако не за друго, то поради следната причина: произвеждащият, ако ще трябва да произведе тоталността на явяването на света, ще трябва да бъде извън него и съответно да бъде нещо неизвестно".

10. Дебатът

Втората част на диалога е размяна на реплики, или дебат между Пенроуз и Северино, а също между публиката и лекторите. Много от въпросите се отнасят до микротубулите. Що се отнася до способността им да преживеят смъртта и да запазят част от спомените от отминалия живот, Пенроуз е твърде скептичен: „Мисля, че микротубулите могат да преживеят смъртта в същата степен като невроните." Втора група въпроси са свързани със съзнанието, в частност с начините да разберем как му въздейства общата анестезия. Споделеното мнение е, че микротубулите имат пряка връзка с действието на общата анестезия. По този повод Пенроуз обсъжда ролята на малкия мозък, който контролира „автоматичните" процеси и чието действие е изцяло несъзнателно, противопоставяйки го на крайния мозък (*cerebrum*), в който осъзнаването играе забележима роля. В потвърждение на идеите си Пенроуз посочва, че микротубулите се срещат в голямо количество в пирамидалните клетки, които на свой ред изобилстват в крайния мозък, макар да не се срещат в малкия мозък. Интересното тук е, че съзнанието, изглежда, е свързано с пирамидалните клетки. Различни тестове, чрез които се изследва осъзнатото разбиране, се осъществяват чрез шахматни задачи, решавани от хора и от компютри. По повод на тях установяваме разликата между „разбирането", например за какво служи една верига от пешки, и чистото „механично" изчисление, което извършва компютърът. Тези тестове ясно показват разликата между осъзнатото мислене (или съзнателното разбиране) и простото изчисление.

11. Съзнание при животните

Според Пенроуз идеята за креативност е заблуждаваща и двусмислена: креативността не е добро мерило за наличие на съзнание. Разбирането е това, което ни позволява да установим разликата между осъзнати и несъзнателни действия. Освен това

много трудно е да различим креативността от случайното създаване на нещо (схващано като различно от това, което е било „създадено“ до момента). Ето защо креативността, за разлика от разбирането, е изключително трудно да бъде потвърдена експериментално и да я оценим обективно. Накрая Пенроуз поддържа идеята, че феноменът на съзнанието присъства и при животните. Това се съгласува с разбирането му, че микротубулите са мястото, от което извират протосъзнанието и в крайна сметка съзнанието. Микротубули имат много от „висшите“ животни, поради което те са надарени със съзнание – това се отнася в частност за кучетата, кравите, слоновете, маймуните, горилите, делфините, мишките и т.н. От това могат да бъдат направени етически изводи, например що се отнася до дължимото зачитане не просто на другите хора, а също и на много от другите живи същества.

12. Съзнание и език

Идеята на Пенроуз и Хамероф (според които съзнанието възниква в резултат от натрупване на елементарни „протосъзнателни“ процеси, водещи в крайна сметка до чудото на човешкия ум) ни води към един смущаващ извод, свързан със „силния“ изкуствен интелект. Както е добре известно и както не за пръв се случва в историята, това е проект, който изхожда от формалните езици и цели да изгради „интелект“ посредством софтуер (т.е. чрез компютърни програми, основани на тези езици), следвайки йерархичен процес от горе надолу. Това се прави в светлината на явната убеденост, че конструирането на интелект може да ни отведе по този път до конструиране на съзнание.

От друга страна, според Пенроуз природата първо конструира елементарни форми на съзнание, които, изглежда, са доста често срещани поне при „висшите“ животни. Едва след това тя създава езици (включително сложни езици), а езиците, поне силно развитите, може би са характерни за само един биологичен вид – за нас, хората.

От тази гледна точка програмата на силния изкуствен интелект изглежда „изкуствена“, в смисъл че се движи в посока, обратна на тази, която следва естествената еволюция. Хората се опитват да създадат съзнание, тръгвайки от езика, докато природата създава езици, изхождайки от (прото)съзнание!

Както веднъж посочи Дъглас Хофстатър, вероятно изкуственият интелект следва да бъде сравняван със съвременен реактивен самолет: постига висока производителност при решаване на конкретен тип задачи, но като цяло е неспособен на куп неща, които една лястовица (аналогът на човешкото съзнание) постига с лекота. Реактивният самолет може да измине разстоянието от Лондон до Милано за час, което е непосилно за лястовица или пингвин. Опитайте обаче да приземите реактивен самолет върху улук на покрива...

13. Мястото на съзнанието?

Отговорът на Северино изявява различията между неговия подход и този на Пенроуз. Северино открито критикува Пенроуз за това, че пренебрегва казаното от него относно „явяването на света като първична форма на съзнание“, макар тази идея да се съдържа, поне според Северино, в произведенията на редица влиятелни мислители като Декарт, Кант и Брауер. Пенроуз, търсейки упорито „мястото на съзнанието“, демонстрира, от гледна точка на Северино, че разбира съзнанието като едно от нещата в света, като аспект на „явяването на света“, което в действителност е първичната форма на съзнателност.

Последната тема, във връзка с която разногласията стават особено ясни, е свързана с практическата страна на науката, с тезата на Северино, че концептуалните артикулации на научното знание позволяват установяване на контрол над света, който превъзхожда постижимото чрез други концептуални артикулации, например чрез връзката със свещеното, т.е. молитвата. Това е причината концептуалната артикулация на съвременното научно знание да е така страховита! Днес тя е извор на най-голямата

мощ в света. Мощта обаче е едно, а истината – съвсем друго. В тази връзка Северино се връща към първоначалния си аргумент относно технологичната мощ на научните теории. Той е убеден, че научните теории в крайна сметка се оценяват според технологичните им възможности за трансформиране на света, а не според способността им да го представят правдиво или да го обяснят ефективно. Тук Северино ни връща към темата за интерсубективния характер на науката. Според Попър, за да има власт, тя трябва да бъде интерсубективно призната, а това означава, че и други трябва да приемат трансформативния потенциал на науката.

14. Наука и технология

Последният коментар на Пенроуз в тази връзка (който според мен е правилен) преутвърждава традиционното разделение между науката и технологията. На този етап в дискусията са засегнати дълбоки философски въпроси. По думите на Пенроуз: „Аз разглеждам науката като опит да открием истини за света. Тя няма връзка с морални проблеми. Имам предвид, че въпросите на морала са отделни от тези на науката." Науката се опитва да разбере как функционира светът. След това идва технологията, която се стреми да приложи научните знания. Технологията има тясна връзка с науката, но не е наука. Технологията има огромно и трайно въздействие върху всекидневието на хората. Това, разбира се, поставя редица морални въпроси. Ето защо според Пенроуз използването на технологиите и в частност добрите и лошите приложения на науката имат тясна връзка с етиката. Пенроуз е еднозначен по този въпрос: „Когато се занимавам с наука, аз се опитвам да развия разбирането ни за начините, по които функционира светът, а не се стремя към власт." Науката не е опит светът да бъде овладян. Това е целта на технологията. Технологията и етиката са области, които са ясно разграничени от науката, макар да зависят от нея – промени ли се науката, те трябва да отчетат това и да разберат какви са последиците за тях от тези промени.

Накрая Пенроуз посочва отново, че идеите, които развива заедно с Хамероф, представят съзнанието като феномен, който не се ограничава до човешките същества, а се проявява и при други животни. Следователно отношенията ни с животните имат морално значение и това е допълнителен пример за факта, че науката влияе на моралните ни убеждения.

15. Чатботове

Текстът на бившата ученичка на Северино – Инес Тестони, поставя няколко интересни въпроса, част от които ще обсъдя тук. В началото на статията е представен експеримент, проведен през 2017 г. от групата на „Фейсбук“ за изследване на изкуствения интелект (*FAIR*). Две компютърни програми са обучени да водят разговор на английски език, след което са оставени да разговарят една с друга на английски, но и по нечовешки начин. Двата чатбота, изглежда, постепенно развиват език, който е непонятен за хората. Диалогът, който те водят, може да се разбира като проява на автономно съзнание в компютрите. Вероятно – провокира ни Тестони – системите с изкуствен интелект предвещават или дори реализират „квантовата машина на Тюринг“, която хората все още не знаят как да конструират.

Тази интерпретация се съгласува чудесно с възгледите на Пътнам [17] и Чалмърс [18], които твърдят, че материалното устройство на ума няма никакво отношение към производството на мисли и съзнанието.

Иначе казано, менталните свойства са организационно инвариантни, в смисъл че материалната им основа може да се променя. Ако дадено ментално състояние е организационно инвариантно, то щом мозъкът умре, е теоретично възможно да заменим сивото вещество със система с изкуствен интелект. Тази вълнуваща перспектива е обсъдена от Дъглас Хофстатър и Даниъл Денет в прочутата им книга „Азът на ума“, а също, макар и в различен контекст, от Уоли Пфистър в прекрасния му, но будещ тревога филм „Превъзходство“.

16. Квантови машини на Тюринг

Както е известно, Роджър Пенроуз категорично отхвърля тезиса на „силния изкуствен интелект“: системите с изкуствен интелект функционират на базата на формален език, докато човешкото мислене е несводимо до изчислителни процеси. Ако сърцевината на (човешкото) съзнание е съставена от неизчислими, неалгоритмични процеси на квантов колапс, протичащи в микротубулите, от това следва, че (човешкото) съзнание не е представимо от обичайна машина на Тюринг, а човешкият ум има способности, които нито една система с изкуствен интелект не може да притежава поради физичните аспекти на неизчислимия *OrchOR* механизъм [19].

Единственият начин, по който поне на теория е възможно да променим това и да се сдобием с възможност да „демонстрираме“ съзнателен процес, е, ако самата машина на Тюринг се основава на квантовомеханични принципи, превърне се в квантова машина на Тюринг. Въз основа на това Тестони формулира следната вълнуваща теза: квантовата механика може би е свързващото звено между органичната и неорганичната материя, което е в основата на ума и съзнанието. Ако е така, то не можем да кажем, че съзнанието е присъщо само на хората точно поради това, че се основава на квантовата механика. Тъкмо обратното – съзнанието по принцип може да бъде манифестирано както от органична, така и от неорганична материя, тъй като квантовата механика е в основата на всички известни форми на материята.

В заключение Тестони посочва, че – според Северино – „съзнанието е феноменална проява на всичко, което е вечно като всяка материя и е тъждествено на самото себе си, така че е несводимо до материята. Връзките между съзнанието и материята (сивото вещество в мозъка или друг вид материя) не могат да се сведат до реципрочна идентичност“. Ако съзнанието (проявата) не се ограничава до черепната кухина, тогава трябва да признаем, че то трансцендира индивидуалното и дори човешкото, съответно способността ни да разпознаваме неговото наличие. Във връзка с това е прокаран паралел с идеята на Се-

верино за съзнанието изобщо като „явяване на света“ и идеята на Пенроуз–Хамероф за „елементи на протосъзнанието“, които би трябвало да са налице навсякъде, където протича колапс на вълновата функция.

17. Машини, които грешат

Текстът на Джузепе Витиело представя най-важните аспекти на дисипативния квантов модел на мозъка (и съзнанието) в квантовата теория на полето. Още в заглавието му Витиело припомня фундаменталното значение на хаоса (дисипативния) за способността на мозъка да реагира гъвкаво на външния свят. Фрийман подчертава значението на тази идея, а бележката, приписвана на Аристотел, според която „мозъкът не е глупава звезда“, нагледно свидетелства, че неговата траектория никога не преминава през една и съща точка по предвидим начин. Мозъкът се държи като „машина, която греши“ – иначе казано, като същностно погрешимо устройство.

„Кохерентност“ е ключовото за подхода на Витиело понятие, чрез което той разбулва чудесата на (човешкия) мозък. Според данните на наблюдението невронната активност в неокортекса свидетелства за формиране на обхватни конфигурации от осцилаторни процеси. Тези конфигурации обхващат области с размери до 20 см от човешкия мозък и практически нулева дисперсия. Наличието на някакъв тип „сътрудничество“ в подобни обхватни области показва, че мозъчната функция не може да бъде обяснена единствено въз основа на знанията ни за отделните елементарни градивни единици на мозъка. Мозъчната активност изисква въвеждане на понятие за „кохерентност“: широко сътрудничество между голям брой неврони в обширни зони от мозъка. Математически инструмент, чрез който може да бъде описана кохерентността, е квантовата теория на полето (КТП). Нейният формализъм е изключително полезен за изследване на биологични системи изобщо и на мозъка в частност. Математическите похвати, описващи „кохерентността“, придават ясно де-

финирано значение на идеята, че макроскопични свойства могат да се основават на микроскопичен динамичен процес. Макроскопичната система притежава физични свойства, каквито не се наблюдават на микроскопско ниво.

Витиело подчертава, че е необходимо да използваме КТП, а не квантовата механика (КМ), тъй като КТП позволява да бъдат описани различни фазови състояния на системата. На практика това се случва, тъй като в КТП има безкрайно много унитарно нееквивалентни представяния на каноничните комутационни съотношения (*CCR*), докато в КМ са допустими само унитарно (съответно физично) еквивалентни описания (за системи с краен брой степени на свобода). Системи, които подобно на мозъка могат да имат различни фазови състояния, трябва да се описват чрез КТП, която отчита множествеността им и преходите между тях – нещо, на което КМ не е способна.

18. Дисипативният квантов мозък

Друго фундаментално значимо наблюдение на Витиело е, че мозъкът е отворена система, която взаимодейства с околната среда. Неговата структурна откритост води Витиело и други учени към формулиране на дисипативен квантов модел на мозъка.

Ключов аспект на този модел е идентифицирането на формирането на „нов", специфичен тип памет с характерно за нея фундаментално състояние (сред безброй много, унитарно нееквивалентни състояния), наречено „вакуум", до което системата мозък – среда има достъп. Според квантовия дисипативен модел това е „тайната" на паметта. Наборът от състояния на паметта, т.е. от състояния на вакуума, може да се опише като „атракторен пейзаж". Мислите се схващат като описващи хаотични траектории, прекосявайки този пейзаж. Всеки акт на разпознаване, на асоцииране с конкретен спомен, може да се представи в рамките на този подход като доближаване до атрактор и последващо улавяне на траекторията от него. Според Витиело това отговаря на интуитивно узнаване, разпознаване на колективно кохерентно

състояние, което по природата си е „неизчислимо“ и непреводимо в логическата рамка на даден език. Фундаменталната „неизчислима“ природа на мисълта се наблюдава и тук, както и във възгледите на Роджър Пенроуз.

Траекториите в пейзажа, изпълнен с атрактори, прескачанията от един спомен към друг, са класически пример за хаотични траектории, макар да свързват квантови състояния. Те не са периодични (никоя траектория не се пресича със самата себе си), а траекториите с еднакви начални състояния също никога не се пресичат – вместо това се разбягват (експоненциално). Тук изходната интуиция на Фрийман намира строг израз: хаотичните траектории са в основата на способността на мозъка да реагира гъвкаво на външния свят, да генерира нови схеми на активация, включително такива, които възприемаме като нови идеи. Този тип „скитане“ е характерна черта на мозъчната активност, на мисленето. Това е причината мозъкът да не е „глупава звезда“. Вместо това той се дължи като „допускащо грешки устройство“, като „машина, която греши“.

Според Витиело неподлежащата на изчислително моделиране активност на това погрешимо устройство е отличителен белег на всяко интелигентно устройство: точно онова, което според Пенроуз е характерна черта на съзнанието.

Идеята, че съзнателният агент по същността си е погрешимо устройство, машина, допускаща грешки, очевидно е несъвместима със стандартната програма на (силния) изкуствен интелект. Тя ни дава точно обратното – покорни, предсказуеми, верни машини, които може би са полезни за подобряване на ограничените ни способности. По същество проектите в областта на изкуствения интелект днес се ограничават до проектиране на „глупави звезди“.

19. Вътрешен опит

Целта на текста на Мауро Д’Ариано и Федерико Фаджин е много амбициозна: авторите представят основните положения на една

теория на „вътрешния опит“ в рамките на квантовата теория на информацията. По същество те предлагат вътрешно съгласувана теория, която решава това, което Дейвид Чалмърс нарича „труден проблем за съзнанието“, а именно въпроса за произхода и свойствата на „вътрешния опит“, който всеки от нас има. Произходът на квалиите (чувствата и усещанията), както и на самия Аз, намира естественото си място в теоретичната схема, обсъдена тук.

Някои от изходните положения в текста са изложени отчасти в други публикации и доклади на Фаджин и Д'Ариано.

Авторите приемат например, точно като Пенроуз, че „истинската“ интелигентност изисква съзнание – нещо, което нашите дигитални компютри нямат и никога няма да придобият. Авторите се противопоставят, подобно на Пенроуз, на стандартното за теорията на изкуствения интелект разбиране, че човешките същества не са нищо повече от органичен хардуер (*wetware*). Те атакуват силната версия на теорията за изкуствения интелект, според която съзнанието е продукт на мозъка и играе роля, аналогична на софтуера в нашите компютри, както и на физикалисткия възглед, че съзнанието „е продукт на функционирането“, че е някакъв вид биологично свойство на живите същества.

Напротив, според авторите същностното свойство на съзнанието е способността за усещане. Разбира се, от способността за усещане следва съществуването на субект, който усеща, на Аз. Ето защо съзнанието е неразривно свързано с Аза, който усеща „вътрешния опит“. Централно място в обсъждането има излагането на теория за „квалиите“ и решението на „трудния проблем за съзнанието“, което обяснява съществуването и динамиката на квалиите.

Авторите възприемат гледната точка на Дейвид Чалмърс, който твърди, че съзнанието е фундаментално свойство, онтологически независимо от всяко известно (или възможно) физично свойство. Всички информационни системи са способни да бъдат съзнателни, съзнанието е несводим аспект или свойство в природата, а не епифеномен. То не е емерджентно свойство. В из-

вестна степен съзнанието е аспект на действителността *ab initio*. Това разбиране води авторите (подобно на Чалмърс) към специфична версия на панпсихизма. Тук са налице явни прилики с идеята на Пенроуз–Хамероф за „атоми“ на протосъзнанието, които трябва да са налични винаги, когато протича колапс на вълновата функция (още повече при „явяването на света“, за което говори Северино).

20. Квантов панпсихизъм

За да решат трудният проблем за съзнанието, за да обяснят нашия опит – сетивен, телесен, умствен и емоционален, включително потока на мислите ни, Д'Ариано и Фаджин се фокусират върху панпсихизма, според който съзнанието е фундаментален аспект на „информацията“, а цялата физика е производна на нея. Според теорията, формулирана тук, фундаментално свойство на информацията е това „да бъде възприемана“ от съхраняваща я „система“.

Информацията, до която се отнася съзнанието, по необходимост е квантова по две причини: присъщата ѝ несподелимост и способността ѝ да изгражда мисли чрез сплитане на квалийни състояния. Авторите определят своя възглед като „панпсихизъм, основан на квантова информация“.

Следвайки Чалмърс, те твърдят, че „съзнанието е фундаментална същност, която не може да бъде обяснена чрез нещо по-просто“. Те обаче не спират дотук, а съотнасят на общите тези на Чалмърс експлицитно формулиран модел и постулират, че „съзнанието е начинът, по който информационната система възприема своето информационно състояние и начина, по който го обработва“. Освен това те постулират, че опитът има квантов характер: информационната теория на съзнанието е квантова теория. След това те излагат основната теза на тяхната теория на квалиите: „Опитът е изграден от структурирани квалии.“ Квалиите (феноменалните качествени определения) са градивни единици на съзнателния опит.