

HOGYAN HOZZUNK LÉTRE EGY ELMÉT

AZ EMBERI GONDOLKODÁS TITKA FELTÁRUL

RAY KURZWEIL

TARTALOM

BEVEZETÉS

1. GONDOLATKÍSÉRLETEK A VILÁGRÓL
 2. GONDOLATKÍSÉRLETEK A GONDOLKODÁSRÓL
 3. A MODELL _____ A _____ NEOCORTEX: _____ A
MINTÁZATFELISMERŐ ELMÉLET
 4. A BIOLÓGIAI NEOKORTEX
 5. AZ ÖREG AGY
 6. TRANSZCENDENS KÉPESSÉGEK
 7. A BIOLÓGIAI IHLETÉSŰ DIGITÁLIS NEOKORTEX
 8. AZ ELME MINT SZÁMÍTÓGÉP
 9. GONDOLATKÍSÉRLETEK AZ ELMÉVEL KAPCSOLATBAN
 10. VISSZATÉRÉS AZ AGYRA ALKALMAZVA
10. KIFOGÁSOK
UTÓSÓ

MEGJEG
YZÉSEK
INDEX

BEVEZETÉS

Az Agy szélesebb, mint az Ég, mert egymás mellé tesszük őket...

Az egyiket a másik könnyedén befogadja - és te - az agy mélyebb, mint a tenger - mert - tartsd meg őket - két a késsel - az egyiket a másik elnyeli - a másik elnyeli...

Ahogy a szivacsok- vödrök-csinálják

Az Agy csak az Isten súlya - Mert - kilóról kilóra - És különbözni fognak - ha különböznek...

Mint szótag a hangból

-Emily Dickinson

A világegyetem legfontosabb jelenségeként az intelligencia képes túllépni a természetes korlátokon, és a világot a saját képére alakítani. Emberi kézben az intelligencia lehetővé tette számunkra, hogy legyőzzük biológiai örökségünk korlátait, és eközben megváltoztassuk önmagunkat. Mi vagyunk az egyetlen faj, amely képes erre.

Az emberi intelligencia története egy olyan világegyetemmel kezdődik, amely képes az információ kódolására. Ez volt az a tényező, amely lehetővé tette az evolúciót. Hogy a világegyetem hogyan vált ilyenné, az maga is egy érdekes történet. A fizika standard modellje több tucatnyi állandót tartalmaz, amelyeknek pontosan annak kell lenniük, amik, különben az atomok nem lettek volna lehetségesek, és nem lettek volna csillagok, bolygók, agyak és agyakra szóló könyvek. Az, hogy a fizika törvényei olyan pontosan vannak beállítva, hogy lehetővé tették az információ evolúcióját, hihetetlenül valószínűtlennek tűnik. Mégis, az antropikus elv alapján nem beszélünk róla, ha nem így lenne. Míg egyesek isteni kezet látnak, addig mások egy multiverzumot látnak, amely az unalmas (információt nem hordozó) univerzumok evolúcióját hozza létre, és az unalmasak kihalnak. De függetlenül attól, hogy a mi univerzumunk hogyan lett olyan, amilyen, a történetünket egy információn alapuló világgal kezdhetjük.

Az evolúció története az absztrakció egyre magasabb szintjein bontakozik ki. Az atomok - különösen a szénatomok, amelyek négy különböző irányú összekapcsolódással gazdag információs struktúrákat képesek létrehozni - egyre összetettebb molekulákat alkottak. Ennek eredményeképpen a fizikából alakult ki a kémia.

Egymilliárd évvel később kialakult a DNS nevű összetett molekula, amely képes volt hosszú információsorokat pontosan kódolni, és e "programok" által leírt szervezeteket létrehozni. Ennek eredményeképpen a kémia szülte a biológiát.

Az organizmusok egyre gyorsabb ütemben fejlesztették ki az idegrendszereknek nevezett kommunikációs és döntési hálózatokat, amelyek képesek voltak koordinálni testük egyre összetettebb részeit, valamint a túlélésüket elősegítő viselkedésformákat. Az

idegrendszereket alkotó idegsejtek egyre intelligensebb viselkedésre képesek agyakká tömörültek. Ily módon a biológiából született meg a neurológia, mivel az agyak immár az információ tárolásának és manipulálásának élvonalába kerültek. Így jutottunk el az atomoktól a molekulákig, a DNS-től az agyáig. A következő lépés egyedülállóan emberi volt.

Az emlősök agya olyan különleges képességgel rendelkezik, amely egyetlen más állatosztályban sem található meg. Képesek vagyunk a *hierarchikus* gondolkodásra, arra, hogy megértsünk egy mintázatba rendezett különböző elemekből álló struktúrát, ezt az elrendezést szimbólummal ábrázoljuk, majd ezt a szimbólumot egy még bonyolultabb elrendezés elemeként használjuk. Ez a képesség a neokortexnek nevezett agyi struktúrában valósul meg, amely az embernél elérte azt a fejlettségi és kapacitásbeli küszöböt, hogy ezeket a mintákat *ötleteknek* nevezzük. Egy véget nem érő rekurzív folyamat révén képesek vagyunk egyre összetettebb eszméket felépíteni. A rekurzívan összekapcsolt ötletek e hatalmas tárházát *tudásnak* nevezzük. Csak a *Homo sapiens* rendelkezik olyan tudásbázissal, amely maga is fejlődik, exponenciálisan növekszik, és generációról generációra öröklődik.

Az agyunk az absztrakció egy újabb szintjét hozta létre, mivel az agyunk intelligenciáját és egy másik lehetővé tevő tényezőt, egy szembefordítható függelék - a hüvelykujjat - használtuk fel arra, hogy a környezetet manipulálva eszközöket építsünk. Ezek az eszközök az evolúció egy új formáját jelentették, mivel a neurológia hozta létre a technológiát. Csak az eszközeinknek köszönhető, hogy tudásbázisunk korlátlanul növekedhetett.

Az első találmányunk a történet volt: a beszélt nyelv, amely lehetővé tette számunkra, hogy a gondolatokat különálló kifejezésekkel ábrázoljuk. Az írott nyelv későbbi feltalálásával különálló formákat fejlesztettünk ki gondolataink szimbolizálására. Az írott nyelv könyvtárai jelentősen kibővítették a segédeszköz nélküli agyunk azon képességét, hogy megőrizze és bővítse a rekurzívan strukturált eszmékből álló tudásbázisunkat.

Vita folyik arról, hogy más fajok, például a csimpánzok képesek-e hierarchikus gondolatokat kifejezni a nyelvben. A csimpánzok képesek megtanulni egy korlátozott számú jelnyelvi szimbólumot, amelyek segítségével kommunikálni tudnak az emberi oktatókkal. Világos azonban, hogy a csimpánzok által kezelni képes tudásstruktúrák összetettségének határozott határai vannak. Az általuk kifejezhető mondatok bizonyos egyszerű főnév-ige szekvenciákra korlátozódnak, és nem képesek az emberekre jellemző komplexitás korlátlan bővítésére. Az ember által létrehozott nyelv bonyolultságának szórakoztató példájához elég csak elolvasni egy látványos, több oldalas mondatokat tartalmazó Gabriel García Márquez-történetet vagy regényt - a hatoldalal "A szellem utolsó útja" című novellája egyetlen mondatból áll, és nagyon jól működik benne

spanyolul és angolul is.¹

A technológiáról szóló három korábbi könyvem (*Az intelligens gépek kora*, amelyet az 1980-as években írtam és 1989-ben adtam ki; *A spirituális gépek kora*, amelyet az 1990-es évek közepén-végén írtam és 1999-ben adtam ki; és *A szingularitás közel van, amelyet a*

2000-es évek elején írtam és 2005-ben adtam ki) fő gondolata az, hogy egy evolúciós folyamat természeténél fogva felgyorsul (az absztrakciós szintek növekedése következtében), és termékeinek komplexitása és képessége exponenciálisan növekszik. Ezt a jelenséget a gyorsuló megtérülés törvényének (LOAR) nevezem, és mind a biológiai, mind a technológiai evolúcióra vonatkozik. A LOAR legdrámaibb példája az információs technológiák kapacitásának és ár/teljesítmény arányának figyelemre méltóan kiszámítható exponenciális növekedése. A technológia evolúciós folyamata változatlanul a számítógéphez vezetett, amely viszont lehetővé tette tudásbázisunk hatalmas bővülését, és kiterjedt kapcsolatokat tett lehetővé egyik tudásterületről a másikra. A web maga is erőteljes és talált példája annak, hogy egy hierarchikus rendszer képes a tudás széles skáláját felölelni, miközben megőrzi a benne rejlő struktúrát. Maga a világ is eredendően hierarchikus - a fák ágakat tartalmaznak, az ágak leveleket, a levelek pedig ereket. Az épületek emeletekből állnak; az emeletek szobákból állnak; a szobák ajtókat, ablakokat, falakat és padlót tartalmaznak.

Olyan eszközöket is kifejlesztettünk, amelyek lehetővé teszik számunkra, hogy saját biológiánkat pontos információkkal megértsük. Gyorsan visszafejtjük a biológia alapjául szolgáló információs folyamatokat, beleértve az agyunkat is. Az emberi genom formájában ma már birtokunkban van az élet tárgyi kódja, ami önmagában is az exponenciális növekedés kiemelkedő példája, hiszen a világ által szekvenált genetikai adatok mennyisége az elmúlt húsz évben évente körülbelül megduplázódott.² Ma már képesek vagyunk számítógépen szimulálni, hogy a bázispárok szekvenciái hogyan eredményezik az aminosavak szekvenciáit, amelyek háromdimenziós fehérjékké állnak össze, amelyekből az egész biológia felépül. A számítási erőforrások exponenciális növekedésével folyamatosan növekszik azoknak a fehérjéknek a bonyolultsága, amelyek esetében a fehérjék összecsukódását szimulálni tudjuk.³ Azt is szimulálni tudjuk, hogy a fehérjék hogyan lépnek kölcsönhatásba egymással az atomi erők bonyolult háromdimenziós táncában. A biológia növekvő megértése az egyik fontos aspektusa annak, hogy felfedezzük az evolúció által ránk bízott intelligens titkokat, majd ezeket a biológiailag inspirált paradigmákat felhasználva egyre intelligensebb technológiát hozunk létre.

Jelenleg egy nagyszabású projekt van folyamatban, amelyben sok ezer tudós és mérnök dolgozik azon, hogy megértsük az intelligens folyamatok legjobb példáját: az emberi agyat. Ez vitathatatlanul az ember-gép civilizáció történetének legfontosabb erőfeszítése. A *The Singularity Is Near (A szingularitás közel van)* című könyvemben kifejtettem, hogy a gyorsuló megtérülés törvényének egyik következménye, hogy más intelligens fajok valószínűleg nem fognak létezni. Összefoglalva az érvelést, ha léteznének, már észrevettük volna őket, tekintve a viszonylag rövid időt, amely eltelt aközött, hogy egy civilizáció nyers technológiával rendelkezett (gondoljunk csak arra, hogy az országos információküldés 1850leggyorsabb módja a Pony Expressz volt), és aközött, hogy olyan technológiával rendelkezett, amely

képes túllépni saját bolygóján.⁴ Ebből a szempontból az emberi agy visszafejtése a világegyetem legfontosabb projektjének tekinthető.

A projekt célja, hogy pontosan megértsük, hogyan működik az emberi agy, majd ezeket a feltárt módszereket felhasználva jobban megértsük magunkat, szükség esetén megjavítsuk az agyat, és - ami e könyv témája szempontjából a legfontosabb - még intelligensebb gépeket hozunk létre. Ne feledjük, hogy egy természeti jelenség nagymértékű felerősítése pontosan az, amire a mérnöki tudomány képes. Példaként gondoljunk a Bernoulli-elv meglehetősen finom jelenségére, amely kimondja, hogy egy mozgó görbe felület fölött valamivel kisebb a légnyomás, mint egy mozgó sík felület fölött. A tudósok még mindig nem tisztázták teljesen, hogy a Bernoulli-elv hogyan hozza létre a szárnyak felhajtóerejét, de a mérnöki tudomány mégis fogta ezt a finom felismerést, összpontosította erejét, és megteremtette a repülés egész világát.

Ebben a könyvben bemutatok egy tézist, amelyet az elme mintafelismerési elméletének (PRTM) nevezek, és amely szerintem leírja a neokortex (az agy érzékelésért, emlékezetért és kritikus gondolkodásért felelős régiója) alapvető algoritmusát. Az előttem álló fejezetekben leírom, hogy a legújabb idegtudományi kutatások, valamint a saját gondolatkísérleteink hogyan vezetnek arra a megkerülhetetlen következtetésre, hogy ezt a módszert következetesen alkalmazzák a neokortexben. A PRTM és a LOAR kombinációjának következménye az, hogy képesek leszünk ezeket az elveket úgy megtervezni, hogy saját intelligenciánk erejét jelentősen kibővítsük.

Ez a folyamat már javában zajlik. Több száz olyan feladat és tevékenység van, amely korábban kizárólag az emberi intelligencia hatáskörébe tartozott, és amelyet ma már számítógépek végezhetnek, általában nagyobb pontossággal és sokkal nagyobb léptékben. Minden alkalommal, amikor e-mailt küldünk vagy mobiltelefon-hívást kapcsolunk, intelligens algoritmusok optimálisan továbbítják az információt. Szerezzen be egy elektrokardiogramot, és olyan számítógépes diagnózist kap, amely vetekszik az orvosokéval. Ugyanez igaz a vérsajtóképekre is. Az intelligens algoritmusok automatikusan felderítik a hitelkártyacsalásokat, repülnek és landolnak repülőgépekkel, irányítják az intelligens fegyverrendszereket, intelligens számítógépes tervezéssel segítik a termékek tervezését, nyomon követik a just-in-time készletszinteket, robotgyárakban szerelik össze a termékeket, és mesteri szinten játszanak olyan játékokat, mint a sakk vagy akár a finom Go.

Emberek milliói látták, ahogy az IBM Watson nevű számítógépe a *Jeopardy!* természetes nyelvű játékában magasabb pontszámot ért el, mint a világ két legjobb emberi játékosa együttvéve. Meg kell jegyezni, hogy Watson nemcsak a *Jeopardy!* kérdésében szereplő finom nyelvezetet (amely olyan jelenségeket tartalmaz, mint a szójátékok és metaforák) olvasta és "értette", hanem a válaszadáshoz szükséges tudást több száz millió oldalnyi természetes nyelvű dokumentum, köztük a Wikipédia és más enciklopédiák megértéséből saját maga szerezte. Az emberi szellemi törekvések gyakorlatilag minden területét el kellett sajátítania, beleértve a történelmet, a tudományt, az irodalmat, a művészeteket, a kultúrát és még sok mást. Az IBM most a Nuance Speech Technologies-szal (korábban Kurzweil Computer Products, az első cégem) együttműködve dolgozik a Watson új változatán, amely az orvosi szakirodalmat (lényegében az összes orvosi folyóiratot és a vezető orvosi blogokat) olvassa, hogy a Nuance klinikai nyelvértési technológiáit felhasználva mesteri diagnosztikussá és orvosi tanácsadóvá váljon. Egyes megfigyelők azzal

érveltek, hogy Watson nem igazán "érti" a *Jeopardy!* feladványait vagy az általa elolvasott lexikonokat, mert csak "statisztikai elemzéssel" foglalkozik. Az egyik legfontosabb pont, amit itt leírok, hogy a mesterséges intelligencia területén kialakult matematikai technikák (mint amilyeneket Watson és Siri, az iPhone asszisztens használ) matematikailag nagyon hasonlóak azokhoz a módszerekhez, amelyeket a biológia a neokortex formájában fejlesztett ki. Ha a nyelv és más jelenségek statisztikai elemzéssel történő megértése nem számít valódi megértésnek, akkor az embereknek sincs megértésük.

A Watson képessége, hogy intelligens módon elsajátítsa a természetes nyelvű dokumentumokban található tudást, hamarosan megjelenik egy Önhöz közeli keresőmotorban. Az emberek már most is természetes nyelven beszélnek a telefonjukkal (például a Siri segítségével, amelyhez a Nuance is hozzájárult). Ezek a természetes nyelvű asszisztensek gyorsan egyre intelligensebbek lesznek, ahogy egyre több Watson-szerű módszert használnak, és ahogy maga Watson is tovább fejlődik.

A Google önvezető autói 200 000 mérföldet tettek meg Kalifornia forgalmas városaiban és településein (ez a szám kétségtelenül sokkal magasabb lesz, mire ez a könyv a valódi és virtuális polcokra kerül). A mesterséges intelligenciának számos más példája is van a mai világban, és még rengeteg további van a láthatáron.

A LOAR további példájaként az agyszkenelés térbeli felbontása és az agyról gyűjtött adatok mennyisége évente megduplázódik. Azt is demonstráljuk, hogy ezeket az adatokat az agyi régiók működő modelljeivé és szimulációivá tudjuk alakítani. Sikerült visszafejtenünk a hallókéreg kulcsfontosságú funkcióit, ahol a hangokkal kapcsolatos információkat dolgozunk fel; a látókéregét, ahol a látásunkból származó információkat dolgozunk fel; és a kisagyét, ahol a képzés egy részét végezzük (például a repülő labda elkapását).

Az emberi agy megértésére, modellezésére és szimulálására irányuló projekt élvonalába tartozik az agyi neokortex visszafejtése, ahol a rekurzív hierarchikus gondolkodásunkat végezzük. Az agykéreg, amely az emberi agy százalékát 80 teszi ki, erősen ismétlődő struktúrából áll, ami lehetővé teszi az emberek számára, hogy tetszőlegesen összetett gondolati struktúrákat hozzanak létre.

Az elme mintafelismerési elméletében leírok egy modellt arról, hogy az emberi agy hogyan éri el ezt a kritikus képességet egy nagyon okos, a biológiai evolúció által tervezett struktúra segítségével. Ennek az agykérgi mechanizmusnak vannak olyan részletei, amelyeket még nem értünk teljesen, de eleget tudunk azokról a funkciókról, amelyeket el kell látnia, hogy ennek ellenére olyan algoritmusokat tervezhessünk, amelyek ugyanazt a célt szolgálják. Azzal, hogy elkezdjük megérteni a neokortexet, most abban a helyzetben vagyunk, hogy nagymértékben felerősíthetjük annak erejét, ahogyan a repülés világa is nagymértékben felerősítette a Bernoulli-elv erejét. A neokortex működési elve vitathatatlanul a világ legfontosabb gondolata, mivel képes minden tudás és készség képviselőjére, valamint új tudás létrehozására. Végül is a neokortex az, amely minden regényért, minden dalért, minden festményért, minden tudományos felfedezésért és az emberi gondolkodás sokféle más termékéért felelős.

Az idegtudományok területén nagy szükség van egy olyan elméletre, amely összeköti a naponta jelentkező, rendkívül eltérő és kiterjedt megfigyeléseket. Az egységes elmélet a tudomány minden jelentős területén alapvető követelmény. Az [1. fejezetben](#) leírom, hogyan egyesítette két álmódó a biológiát és a fizikát, olyan területeket, amelyek korábban reménytelenül rendezetlenek és változatosnak tűntek, majd foglalkozom azzal, hogyan alkalmazható egy ilyen elmélet az agy tájékán.

Manapság gyakran találkozunk az emberi agy összetettségének nagy ünneplésével. A Google mintegy 30 millió linket ad vissza az erre a témára vonatkozó idézetekre vonatkozó keresésre. (Ezt azonban lehetetlen lefordítani a ténylegesen visszaküldött idézetek számára, mivel a linkelt weboldalak némelyikén több idézet is szerepel, némelyiken pedig egy sem.) Maga James D. Watson 1992-ben azt írta, hogy "az agy az utolsó és legnagyobb biológiai határ, a legösszetettebb dolog, amit eddig felfedeztünk a világegyetemünkben". A továbbiakban elmagyarázza, hogy miért hiszi, hogy "több százmilliárd sejtet tartalmaz, amelyek egymáshoz több trillió kapcsolaton keresztül kapcsolódnak. Az agy megdöbbeneti a elme."⁵

Egyetértek Watson véleményével, miszerint az agy a legnagyobb biológiai határterület, de az a tény, hogy sok milliárd sejtet és trilliónyi kapcsolatot tartalmaz, nem feltétlenül teszi az elsődleges módszerét összetetté, ha könnyen érthető (és újraalkotható) mintákat tudunk azonosítani ezekben a sejtekben és kapcsolatokban, különösen a tömegesen redundánsakat.

Gondoljuk át, mit jelent összetettnek lenni. Megkérdezhetnénk, hogy egy erdő komplex-e? A válasz attól függ, hogy milyen nézőpontot választunk. Megállapíthatjuk, hogy az erdőben sok ezer fa van, és mindegyik más és más. Megállapíthatjuk azt is, hogy minden fának több ezer ága van, és minden ág teljesen más. Aztán leírhatnád egyetlen ág tekervényes szeszélyeit. A következtetés az lehet, hogy az erdő olyan komplexitással rendelkezik, amely meghaladja legvadabb képzeletünket.

Egy ilyen megközelítés azonban szó szerint azt jelentené, hogy nem látjuk a fától az erdőt. Természetesen a fák és az ágak között rengeteg fraktálvariáció van, de ahhoz, hogy helyesen megértsük az erdő alapelveit, jobban tennénk, ha a sztochasztikus (azaz véletlenszerű) variációval járó redundancia határozott mintáinak azonosításával kezdenénk. Joggal mondhatjuk, hogy az erdő fogalma egyszerűbb, mint a fa fogalma.

Így van ez az agyban is, amely hasonlóan hatalmas redundanciával rendelkezik, különösen a neokortexben. Amint azt ebben a könyvben leírom, joggal mondhatjuk, hogy egyetlen neuronban több komplexitás van, mint a neokortex teljes szerkezetében.

Ebben a könyvben semmiképpen sem az a célom, hogy újabb idézettel egészítsem ki a már létező milliányi idézetet, amelyek tanúsítják, hogy az agy mennyire összetett, hanem inkább az, hogy lenyűgözzem Önt az egyszerűségének erejével. Ezt úgy fogom elérni, hogy leírom, hogy a neokortexben százmilliószor ismétlődő, alapvető, zseniális mechanizmus egy minta felismerésére, megjegyzésére és előrejelzésére hogyan magyarázza gondolkodásunk nagyfokú sokszínűségét. Ahogyan az organizmusok elképesztő sokfélesége a nukleáris és mitokondriális DNS-ben található genetikai kód értékeinek különböző kombinációiból ered, ugyanúgy az eszmék, gondolatok és készségek elképesztő sokasága alakul ki a neokortikális mintafelismerőinkben és a köztük található minták (kapcsolatok és szinaptikus erősségek)

értékei alapján. Ahogy Sebastian Seung, az MIT idegtudósa mondja: "Az identitás nem a génjeinkben, hanem a kapcsolatokban rejlik".

az agysejtjeink között."⁶

Különbséget kell tennünk a tervezés valódi komplexitása és a látszólagos komplexitás között. Gondoljunk csak a híres Mandelbrot-halmazra, amelynek képe már régóta a komplexitás szimbóluma. A látszólagos bonyolultság értékeléséhez hasznos, ha ráközelítünk a képére (amelyet az ebben a végjegyzetben található linkeken keresztül érhet el).⁷ A végtelen

bonyolultság a bonyolultságon belül, és mindig különböznek egymástól. Mégis, a Mandelbrot-halmaz felépítése - a képlet - nem is lehetne egyszerűbb. Hat karakter hosszú: $Z = Z^2 + C$, amelyben Z egy "komplex" szám (azaz egy számpár), C pedig egy konstans. Nem szükséges teljesen megérteni a Mandelbrot-függvényt ahhoz, hogy lássuk, hogy egyszerű.

Ezt a képletet iteratív módon és a hierarchia minden szintjén alkalmazzuk. Ugyanez igaz az agyra is. Ismétlődő szerkezete nem olyan egyszerű, mint a Mandelbrot-halmaz hatjegyű képlete, de közel sem olyan bonyolult, mint azt az agy bonyolultságára vonatkozó milliárdnyi idézet sugallja. Ez a neokortikális felépítés újra és újra megismétlődik a neokortex által képviselt fogalmi hierarchia minden szintjén. Einstein jól megfogalmazta a könyvvel kapcsolatos céljaimat, amikor azt mondta, hogy "minden intelligens bolond képes nagyobbá és bonyolultabbá tenni a dolgokat... de ahhoz... nagy bátorság kell, hogy az ellenkező irányba haladjunk".

A Mandelbrot-halmaz megjelenítésének egyik nézete, egy egyszerű képlet, amelyet iteratíván alkalmaznak. Ahogy ráközelítünk a kijelzőre, a képek folyamatosan változnak, látszólag bonyolult módon.

Eddig az agyról beszéltem. De mi a helyzet az elmével? Például hogyan jut el a problémamegoldó neokortex a tudatossághoz? És ha már itt tartunk, hány tudatos elme van az agyunkban? Vannak bizonyítékok, amelyek arra utalnak, hogy egynél több is lehet.

Az elmével kapcsolatos másik lényeges kérdés az, hogy mi a szabad akarat, és van-e szabad akaratunk? Vannak olyan kísérletek, amelyek azt mutatják, hogy döntéseinket már azelőtt elkezdjük végrehajtani, mielőtt még tudatában lennénk annak, hogy meghoztuk őket. Ez azt jelenti, hogy a szabad akarat illúzió?

Végül, agyunk mely tulajdonságai felelősek az identitásunk kialakításáért? Ugyanez a személy vagyok, aki hat hónappal ezelőtt voltam? Nyilvánvalóan nem vagyok pontosan ugyanaz, mint akkor voltam, de vajon ugyanaz az identitásom?

Áttekintjük, hogy a mintafelismerés elmélete mit jelent ezekkel az ősi kérdésekkel kapcsolatban.

GONDOLATKÍSÉRLETEK A VILÁGRÓL

Darwin természetes szelekcióról szóló elmélete nagyon későn jelent meg a gondolkodás történetében.

Vajon azért késett, mert szembement a kinyilatkoztatott igazsággal, mert teljesen új témát jelentett a tudomány történetében, mert csak az élőlényekre volt jellemző, vagy mert a céllal és a végső okokkal foglalkozott anélkül, hogy a teremtés aktusát tételezte volna? Szerintem nem. Darwin egyszerűen felfedezte a szelekció szerepét, egy olyan ok-okozati összefüggést, amely nagyon különbözik a tudomány addigi push-pull mechanizmusaitól. Az élőlények fantasztikus sokféleségének eredetét azzal lehetett magyarázni, hogy milyen újszerű, esetleg véletlen eredetű tulajdonságok járultak hozzá a túléléshez. A fizikai vagy biológiai tudományokban nemigen volt olyan, vagy nem volt semmi, ami a szelekciót mint oksági elvet előrevetítette volna.

-B. F. Skinner

Vége semmi sem szent, csak a saját elméd integritása.

-Ralph Waldo Emerson

Egy metafora a geológiából

A XIX. század elején a geológusok egy alapvető kérdésen töprengtek. Az olyan nagy barlangok és kanyonok, mint a Grand Canyon az Egyesült Államokban és a görögországi Vikos-szurdok (állítólag a világ legmélyebb kanyonja) szerte a világon léteztek. Hogyan kerültek oda ezek a fenséges képződmények?

Mindig volt egy vízfolyás, amely kihasználta a lehetőséget, hogy átfolyik ezeken a természetes struktúrákon, de a tizenkilencedik század közepe előtt abszurdnak tűnt, hogy ezek a szelíd áramlások ilyen hatalmas völgyeket és sziklákat hozhattak létre. Charles Lyell brit geológus (1797-1875) azonban azt javasolta, hogy valóban a víz mozgása vájta ki ezeket a nagy geológiai átalakulásokat nagy időintervallumokon keresztül, lényegében egy-egy közetszemcsét. Ezt a javaslatot kezdetben nevetség tárgyává tették, de két évtizeden belül Lyell tézise általánosan elfogadottá vált.

Charles Darwin (1809-1882) angol természettudós volt az, aki figyelmesen figyelte a tudományos közösség Lyell radikális tézisére adott reakcióját. Gondoljunk csak a biológia körül kialakult helyzetre A 1850.terület végtelenül összetett volt, számtalan állat- és növényfajjal állt szemben, amelyek bármelyike nagy bonyolultságot mutatott. Ha valami, akkor a legtöbb tudós ellenállt minden olyan kísérletnek, amely a természet káprázatos változatosságának egységesítő elméletét próbálta megalkotni. Ez a sokféleség Isten teremtésének dicsőségét bizonyította, nem is beszélve a tudósok intelligenciájáról, akik képesek voltak azt elsajátítani.

Darwin úgy közelítette meg a fajok általános elméletének kidolgozását, hogy Lyell

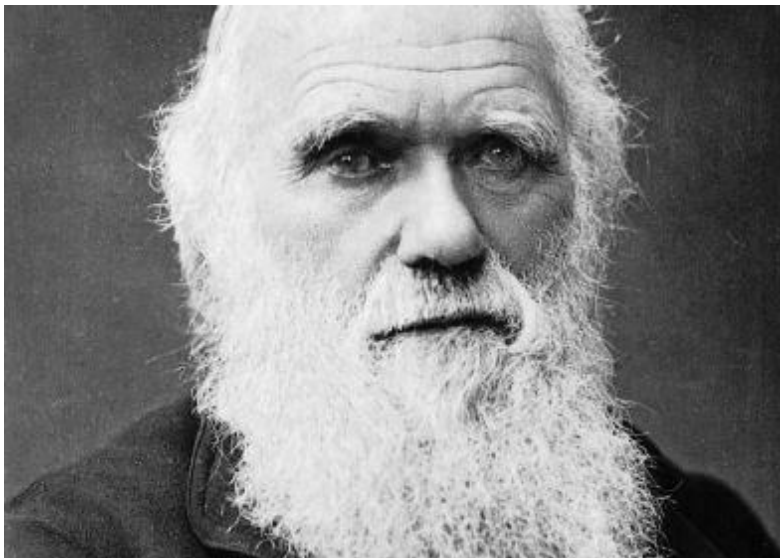
tézisével analóg módon magyarázatot keresett a fajok jellemzőinek sok nemzedék alatt bekövetkező fokozatos változására. Ezt a felismerést saját gondolatkísérleteivel és megfigyeléseivel kombinálta a híres *Beagle-útjában*. Darwin érvelése szerint

hogy minden egyes generációban azok az egyedek, amelyek a legjobban fenn tudtak maradni a saját ökológiai részükben, létrehozzák a következő generációt.

1859. november 22-én került forgalomba Darwin *A fajok eredetéről* című könyve, amelyben világossá tette Lyell iránti adósságát:

Jól tudom, hogy a természetes szelekciónak ezt a tanítását, amelyet a fenti képzeletbeli példák szemléltetnek, ugyanazok az ellenvetések érik, amelyeket kezdetben Sir Charles Lyell nemes nézetei ellen emeltek "a föld modernkori változásai, mint a geológia szemléltető példái" ellen; de ma már nagyon ritkán halljuk, hogy például a parti hullámok hatását jelentéktelen és jelentéktelen oknak nevezik, amikor óriási völgyek feltárására vagy a leghosszabb belvízi sziklák kialakulására alkalmazzák. A természetes szelekció csak végtelenül apró, öröklött módosítások megőrzése és felhalmozódása révén tud hatni, amelyek mindegyike a megőrzött lény számára előnyös; és ahogy a modern geológia szinte száműzte az olyan nézeteket, mint egy nagy völgy egyetlen diluviális hullám által történő kiásása, úgy a természetes szelekció, ha igaz elv, száműzni fogja az új szerves lények folyamatos keletkezéséről vagy bármilyen nagy és hirtelen változásról alkotott hiedelmeket is.

szerkezetük.¹



Charles Darwin, *A fajok eredetéről* című könyv szerzője, amely megalapozta a biológiai evolúció gondolatát.

Mindig több oka is van annak, hogy a nagy új ötleteknek ellenállnak, és Darwin esetében nem nehéz azonosítani ezeket. Az, hogy nem Istentől, hanem a majmoktól, és azelőtt a férgektől származunk, sok kommentátornak nem tetszett. Az a felvetés, hogy a házikutyánk a mi unokatestvérünk, akárcsak a hernyó, nem is beszélve a növényről, amelyen jár (talán a milliommodik vagy milliárdodik unokatestvérünk, de még mindig rokon), sokak számára istenkáromlásnak tűnt.

Az ötlet azonban gyorsan megfogott, mert koherenciát teremtett a korábban látszólag egymástól független megfigyelések sokaságában. 1872-ben, *A fajok eredetéről* szóló könyv hatodik kiadásának megjelenésekor Darwin ezt a szövegrészt tette hozzá: "A dolgok korábbi állásának feljegyzéseként az előző bekezdésekben... több olyan mondatot is megtartottam, amelyek azt sugallják, hogy a természettudósok hisznek az egyes fajok külön-külön teremtésében; és sok kritikát kaptam azért, hogy így fejeztem ki magam. De kétségtelenül ez volt az általános meggyőződés, amikor a jelen mű első kiadása megjelent.

megjelent.... Mostanra a dolgok teljesen megváltoztak, és szinte minden természettudós elismeri az evolúció nagyszerű elvét."²

A következő évszázad során Darwin egységesítő gondolata elmélyült. 1869-ben, mindössze egy évtizeddel A *fajok eredetéről* szóló könyv eredeti megjelenése után, Friedrich Miescher (1844-1895) svájci orvos felfedezett a sejtmagban egy általa "nukleinnek" nevezett anyagot, amelyről kiderült, hogy a DNS.³ 1927-ben Nyikolaj Koltsov (1872-1940) orosz biológus leírta az általa "óriási örökítő molekulának" nevezett anyagot, amely szerinte "két-tükörszázból áll, amelyek félig konzervatív módon, mindkét szálat sablonként használva replikálódnak". Az ő megállapítását is sokan elítélték. A kommunisták fasiszta propagandának tartották, és hirtelen, váratlan halálát a Szovjetunió titkosrendőrségének tulajdonították.⁴ 1953-ban, közel egy évszázaddal Darwin korszakalkotó könyvének megjelenése után James D. Watson (szül. 1928) amerikai biológus és Francis Crick (1916-2004) angol biológus először jellemezte pontosan a DNS szerkezetét, két hosszú, csavarodó molekulából álló kettős spirálként írva le.⁵ Érdeemes kiemelni, hogy megállapításuk a ma már "51-es fotónak" nevezett felvételen alapult, amelyet kollégájuk, Rosalind Franklin készített röntgenkrisztallográfiával, és amely az első olyan ábrázolás volt, amely a kettős spirált ábrázolta. Tekintettel a Franklin képéből származó felismerésekre, voltak olyan javaslatok, hogy osztoznia kellett volna Watson és Crick Nobel-díjában.⁶



Rosalind Franklin (röntgenkrisztallográfia segítségével) elkészítette a DNS-ről azt a kritikus képet, amely lehetővé tette Watson és Crick számára, hogy először írják le pontosan a DNS szerkezetét.

Egy olyan molekula leírásával, amely képes volt kódolni a biológia programját, a biológia egységesítő elmélete szilárdan a helyére került. Egyszerű és elegáns alapot biztosított az egész élet számára. Csupán a sejtmagban (és kisebb mértékben a mitokondriumokban) található DNS-szálakat alkotó bázispárok értékeitől függően egy organizmusból fűszál vagy ember válhat. Ez a felismerés nem szüntette meg a természet elragadó sokféleségét, de ma már értjük, hogy a természet rendkívüli sokfélesége az ezen az univerzális molekulán kódolható struktúrák nagy választékából ered.

Lovaglás egy fénysugáron

A huszadik század elején a fizika világa egy újabb gondolatkísérlet-sorozat révén felborult. Egy német mérnöknek és egy háziasszonynak egy 1879fiú született. Hároméves koráig nem kezdett el beszélni, és állítólag már kilencévesen gondjai voltak az iskolában. Tizenhat évesen már arról álmodozott, hogy egy holdsugáron lovagol.

Ez a fiatal fiú ismerte Thomas Young (1773-1829) angol matematikus 1803-as kísérletét, amely megállapította, hogy a fény hullámokból áll. Az akkori következtetés az volt, hogy a fényhullámoknak valamilyen közegen kell áthaladniuk; elvégre az óceán hullámai a vízen, a hanghullámok pedig a levegőn és más anyagokon keresztül terjednek. A tudósok "éternek" nevezték el azt a közegét, amelyen keresztül a fényhullámok haladnak. A fiú ismerte Albert Michelson (1852- 1931) és Edward Morley (1838-1923) amerikai tudósok 1887-es kísérletét is, amely az éter létezését próbálta igazolni. Ez a kísérlet a folyón evezős csónakkal való fel- és lefelé haladás analógiáján alapult. Ha az evezősök meghatározott sebességgel eveznek, akkor a parttól mért sebességük gyorsabb lesz, ha a folyóval együtt eveznek, mintha ellene mennének. Michelson és Morley azt feltételezte, hogy a fény állandó sebességgel (azaz fénysebességgel) halad az éterben. Arra a következtetésre jutottak, hogy a napfény sebessége, amikor a Föld a Nap felé halad a pályáján (a Földön lévő nézőpontunkból mérve), szemben a látszólagos sebességével, amikor a Föld távolodik a Naptól, eltérőnek kell lennie (a Föld sebességének kétszeresével). Ennek bizonyítása megerősítené az éter létezését. Azonban azt fedezték fel, hogy a Földön áthaladó napfény sebessége nem különbözik, függetlenül attól, hogy a Föld hol áll a pályáján. Eredményeik megcáfolták az "éter" gondolatát, de mi is történt valójában? Ez csaknem két évtizedig rejtély maradt.

Ahogy ez a német tinédzser elképzelte, hogy egy fényhullám mellett halad, arra gondolt, hogy a fényhullámokat megdermedve kellene látnia, ugyanúgy, ahogyan egy vonat sem mozog, ha a vonattal azonos sebességgel haladunk mellette. Rájött azonban, hogy ez lehetetlen, mert a fénysebességnek állandónak kell lennie, függetlenül a saját mozgásától. Ezért inkább azt képzelte el, hogy a fénysugár mellett halad, de valamivel lassabb sebességgel. Mi lenne, ha a fénysebesség százalékával90 utazna? Ha a fénysugarak olyanok, mint a vonatok, gondolta, akkor látnia kellene, hogy a fénysugár a fénysebesség százalékával10 halad előtte. Valójában a földi megfigyelők is ezt látnák. De tudjuk, hogy a fénysebesség állandó, amint azt a Michelson-Morley-kísérlet kimutatta. Tehát a fénysugarat szükségszerűen a teljes fénysebességgel látná maga előtt haladni. Ez ellentmondásnak tűnt - hogyan lehetséges ez?

A válasz a német fiú - akit egyébként Albert Einsteinnek (1879-1955) hívtak - számára huszonhat éves korára vált nyilvánvalóvá. Nyilvánvalóan - az ifjú Einstein mester számára maga az idő is lelassult. Érvelését egy 1905-ben megjelent tanulmányában fejtette ki.² Ha a földi megfigyelők a fiatal

ember óráját tízszer lassabban ketyegne. Sőt, amikor visszatérne a Földre, az órája azt mutatná, hogy csak 10 százalékkal kevesebb idő telt el (egyelőre figyelmen kívül hagyva a gyorsulást és a lassulást). Az ő szemszögéből nézve azonban az órája normálisan ketyegett,

és a mellette lévő fénysugár fénysebességgel haladt. Magának az idő sebességének tízszeres lelassulása (a földi órákhoz képest) teljes mértékben megmagyarázza a látszólagos perspektívaeltéréseket. Szélsőséges esetben az idő múlásának lassulása a nullát érné el, amint az utazási sebesség elérné a fénysebességet; ezért lehetetlen volt a fénysugárral együtt utazni. Bár lehetetlen volt a fénysebességgel utazni, kiderült, hogy elméletileg nem lehetetlen a fénysugárnál *gyorsabban* haladni. Az idő ekkor visszafelé haladna.

Ez az állásfoglalás sok korai kritikus számára abszurdnak tűnt. Hogyan lassulhatna le maga az idő, ha csak valakinek a mozgási sebességén alapulna? Valóban, tizennyolc éven át (a Michelson-Morley-kísérlet óta) más gondolkodók képtelenek voltak belátni egy olyan következtetést, amely Einstein mester számára ennyire nyilvánvaló volt. Sokan mások, akik a tizenkilencedik század második felében végiggondolták ezt a problémát, lényegében "leestek a lóról", ami egy elv következményeinek végigkövetését illeti, ehelyett ragaszkodtak az előzetesen kialakított elképzeléseikhez arról, hogy a valóságnak hogyan kell működnie. (Ezt a metaforát valószínűleg "leesett a fénysugárról" metaforára kellene változtatnom.)

Einstein második elmekísérlete az volt, hogy saját magát és testvérét az űrben repülve képzelte el. 186 000 mérföldre vannak egymástól. Einstein gyorsabban akar haladni, de azt is szeretné, ha a köztük lévő távolság változatlan maradna. Ezért minden alkalommal, amikor gyorsítani akar, egy zseblámpával jelez a bátyjának. Mivel tudja, hogy egy másodpercbe telik, amíg a jel eléri a bátyját, egy másodpercet vár (a jel elküldése után), hogy elindítsa a saját gyorsulását. Minden alkalommal, amikor a testvér megkapja a jelet, azonnal gyorsít. Így a két testvér pontosan ugyanabban az időben gyorsul, és ezért állandó távolságban maradnak egymástól.

De most gondoljunk bele, mit látnánk, ha a Földön állnánk. Ha a testvérek távolodnának tőlünk (Albert vezetne), úgy tünne, hogy a fénynek kevesebb mint egy másodpercbe telik, amíg eléri a testvért, mert ő a fény felé halad. Azt is látnánk, hogy Albert testvérenek órája lelassul (mivel a sebessége nő, ahogy közelebb kerül hozzánk). Mindkét ok miatt azt látnánk, hogy a két testvér egyre közelebb kerül egymáshoz, és végül összeütközik. Mégis a két testvér szemszögéből nézve, ők továbbra is egy

állandó mérföldekre 186,000 egymástól.

Hogyan lehetséges ez? A válasz nyilvánvalóan az, hogy a távolságok a mozgással párhuzamosan (de nem merőlegesen) zsugorodnak. Tehát a két Einstein testvér egyre rövidebb lesz (feltételezve, hogy fejfelé előre repülnek), ahogy gyorsulnak. Ez a bizarr következtetés valószínűleg több korai rajongót veszített Einstein számára, mint az idő múlásának különbsége.

Ugyanebben az évben Einstein az anyag és az energia kapcsolatát egy újabb elmekísérlettel vizsgálta. James Clerk Maxwell skót fizikus már az 1850-es években kimutatta, hogy a fotonoknak nevezett fényrészecskének nincs tömegük, de mégis hordoznak impulzusokat. Gyerekkoromban volt egy Crookes sugázmérő nevű készülékem,⁸

amely egy légmentes üveggömbből állt, amely részleges vákuumot tartalmazott, és egy négy szárnyból álló, orsón forgó lapátkészletből állt. A lapátok egyik oldala fehér, a másik fekete volt. Az egyes lapátok fehér oldala visszaverte a fényt, a fekete oldala pedig elnyelte a fényt. (Ezért hűvösebb egy forró napon fehér pólót viselni, mint egy feketét.) Amikor egy fényt ráirányítottak az eszközre, a lapátok elfordultak, a sötét oldaluk pedig eltávolodott a fénytől. Ez közvetlen bizonyítéka annak, hogy a fotonok elég lendületet hordoznak ahhoz, hogy ténylegesen okozzanak

a radiométer lapátjainak mozgását.⁹

Einstein azzal a problémával küzdött, hogy a lendület a tömeg függvénye: A lendület egyenlő a tömeg és a sebesség szorzatával. Így egy mérföld 30 per órával haladó mozdony sokkal nagyobb lendülettel rendelkezik, mint mondjuk egy ugyanilyen sebességgel haladó rovar. Hogyan lehet tehát egy nulla tömegű részecskének pozitív impulzusa?

Einstein elmekísérlete egy térben lebegő dobozból állt. A doboz belsejében egy fotont bocsátanak ki balról a jobb oldal felé. A rendszer teljes impulzusának meg kell maradnia, tehát a doboznak a foton kibocsátása után balra kellene visszahullnia. Egy bizonyos idő elteltével a foton ütközik a doboz jobb oldalával, és impulzusát visszaviszi a dobozba. A rendszer teljes impulzusa ismét megmarad, így a doboz most már nem mozog.



Crookes sugárzásmérő - a négyszárnyú szárnyas lapát forog, amikor fény éri.

Eddig minden rendben. De nézzük meg a perspektívát Einstein úr szemszögéből, aki kívülről figyeli a dobozt. Nem lát semmilyen külső hatást a dobozra: Egyetlen részecske - tömeggel vagy anélkül - sem éri el, és semmi sem hagyja el. A fenti forgatókönyv szerint Einstein úr mégis azt látja, hogy a doboz átmenetileg balra mozog, majd megáll. Elemzésünk szerint minden egyes fotonnak állandóan balra kellene mozdognia a dobozra. Mivel a dobozra vagy a dobozból nem hatott semmilyen külső hatás, a tömegközéppontjának ugyanazon a helyen kell maradnia. A doboz belsejében lévő foton azonban, amely balról jobbra mozog, nem változtathatja meg a tömegközéppontot, mert nincs tömege.

Vagy mégis? Einstein arra a következtetésre jutott, hogy mivel a foton egyértelműen rendelkezik energiával és impulzussal, ezért a fotonnak rendelkeznie kell egy tömegegyenértékkel is. A mozgó foton energiája teljes mértékben egyenértékű egy mozgó tömeggel. Kiszámíthatjuk, hogy mi ez az ekvivalencia, ha felismerjük, hogy a rendszer tömegközéppontjának a foton mozgása során mozdulatlanul kell maradnia. A matematikát kidolgozva Einstein megmutatta, hogy a tömeg és az energia egyenértékű, és egy egyszerű konstanssal függ össze. Volt azonban egy bökkenő: A konstans lehet, hogy egyszerű, de kiderült, hogy óriási: a fénysebesség négyzete (körülbelül $1,7 \times 10^{17}$ méter² másodpercenként² - azaz 17, amelyet 16 nulla követ). Így kapjuk Einstein híres $E = mc^2$

értékét.¹⁰ Így egy unciá (28 gramm) tömeg 600 000 tonna TNT-nek felel meg. Einstein 1939. augusztus 2-án Roosevelt elnöknek írt levele, amelyben tájékoztatta őt az atombomba lehetőségéről, amely ezen a képleten alapul, és amely az atomkorszakot nyitotta meg.¹¹

Azt gondolhatnánk, hogy ennek már korábban is nyilvánvalónak kellett volna lennie, mivel a kísérletezők észrevették, hogy a radioaktív anyagok tömege a sugárzás hatására idővel csökken. Feltételezték azonban, hogy a radioaktív anyagok valamilyen speciális, nagy energiájú üzemanyagot tartalmaznak, amely elég. Ez a feltételezés nem teljesen téves; csakhogy a "kiégő" üzemanyag egyszerűen tömeg volt.

Több oka is van annak, hogy ezt a könyvet Darwin és Einstein elmekísérleteivel nyitottam. Először is, megmutatják az emberi agy rendkívüli erejét. Egy tollon és papíron kívül semmilyen más eszköz nélkül, hogy megrajzolja az egyszerű elmekísérletek pályakaalakjait, és leírja az ezekből következő, meglehetősen egyszerű egyenleteket, Einstein képes volt megdönteni a fizikai világ két évszázados megértését, mélyen befolyásolni a történelem menetét (beleértve a második világháborút is), és bevezetni a nukleáris korszakot.

Igaz, hogy Einstein a XIX. század néhány kísérleti eredményére támaszkodott, bár ezek a kísérletek sem használtak kifinomult berendezéseket. Az is igaz, hogy Einstein elméleteinek későbbi kísérleti igazolása fejlett technológiákat használt, és ha ezeket nem fejlesztették volna ki, akkor nem lenne meg az a mai igazolásunk arról, hogy Einstein elképzelései hitelesek és jelentősek. Ezek a tényezők azonban nem vonnak le semmit abból a tényből, hogy ezek a híres gondolat-kísérletek az emberi gondolkodás erejét mutatják meg a legjobban.

Einsteint széles körben a huszadik század vezető tudósaként tartják számon (Darwin a tizenkilencedik században jó esélyes lett volna erre a megtiszteltetésre), ugyanakkor az elméleteinek alapjául szolgáló matematika végső soron nem túl bonyolult. Maguk a gondolat-kísérletek egyszerűek voltak. Elgondolkozhatunk tehát azon, hogy Einstein milyen szempontból tekinthető különösen okosnak. Később még megbeszéljük, hogy pontosan mit is csinált az agyával, amikor az elméleteivel előállt, és hol rejlik ez a tulajdonsága.

Ezzel szemben ez a történelem az emberi gondolkodás korlátait is megmutatja. Einstein képes volt meglovagolni a fénysugarát anélkül, hogy leesett volna (bár arra a következtetésre jutott, hogy valójában lehetetlen meglovagolni egy fénysugarat), de hány ezer más megfigyelő és gondolkodó volt teljesen képtelen végiggondolni ezeket a figyelemre méltóan egyszerű feladatokat? Az egyik közös kudarc az, hogy a legtöbb embernek nehézséget okoz, hogy elveti és meghaladja társai gondolatait és nézőpontjait. Vannak más elégtelenségek is, amelyeket részletesebben is tárgyalni fogunk, miután megvizsgáltuk, hogyan működik a neokortex.

A neokortex egységes modellje

A legfontosabb ok, amiért megosztom a történelem talán leghíresebb gondolat kísérleteit, az az, hogy bevezetést adjak ahhoz, hogy ugyanezt a megközelítést az agyra vonatkozóan is alkalmazzuk. Mint látni fogják, néhány egyszerű gondolat kísérleten keresztül figyelemre méltóan messzire juthatunk az emberi intelligencia működésének megismerésében. Tekintettel az érintett témára, az elmekísérletek nagyon is megfelelő megközelítésnek kell lenniük.

Ha egy fiatal ember tétlen gondolatai és a tollon és papíron kívül semmilyen más eszköz használata elegendő lenne ahhoz, hogy forradalmasítsa a fizika megértését, akkor egy olyan jelenséggel kapcsolatban, amelyet sokkal jobban ismerünk, ésszerű előrehaladást kellene elérnünk. Elvégre ébrenlétünk minden pillanatában - és álmainkban is - megtapasztaljuk a gondolkodásunkat.

Miután felépítettünk egy modellt arról, hogyan működik a gondolkodás az önreflexió folyamatán keresztül, megvizsgáljuk, milyen mértékben tudjuk ezt megerősíteni a tényleges agyak legújabb megfigyelései és a folyamatok gépi újratermítésének jelenlegi állása alapján.

FEJEZET 2

GONDOLATKÍSÉRLETEK A GONDOLKODÁSRÓL

Nagyon ritkán gondolkodom szavakban. Jön egy gondolat, és lehet, hogy utána megpróbálom szavakkal kifejezni.

-Albert Einstein

Az agy egy háromkilós tömeg, amelyet a kezében tarthatsz, és amely képes elképzelni egy százmilliárd fényév átmérőjű univerzumot.

-Marian Diamond

Megdöbbenőnek tűnik, hogy egy mindössze háromkilós tárgy, amely ugyanazokból az atomokból áll, amelyek minden mást alkotnak a nap alatt, képes gyakorlatilag mindent irányítani, amit az emberek eddig tettek: elrepülni a Holdra és hetven hazafutást ütni, megírni a *Hamletet* és felépíteni a Taj Mahalt - sőt, még magának az agynak a titkait is megfejteni.

-Joel Havemann

Körülbelül 1960, abban az évben kezdtem el gondolkodni a gondolkodásról, amikor felfedeztem a számítógépet. Ma már nehéz lenne olyan tizenkét éves gyereket találni, aki nem használ számítógépet, de akkoriban még csak egy maroknyi volt belőlük a szülővárosomban, New York Cityben. Persze ezek a korai eszközök nem fértek el a kezében, és az első, amelyhez hozzáférhettem, egy nagy szobát foglalt el. Az 1960-as évek elején egy IBM 1620-as számítógépen programoztam, hogy varianciaelemzéseket (statisztikai tesztek) végezzek azokon az adatokon, amelyeket egy kora gyermekkori nevelési program, a Head Start elődjének tanulmányozása során gyűjtöttek. Ezért jelentős drámai volt az erőfeszítés, mivel ennek az országos oktatási kezdeményezésnek a sorsa a mi munkánkon múlt. Az algoritmusok és az elemzett adatok eléggé összetettek voltak ahhoz, hogy nem tudtuk előre megjósolni, milyen válaszokat fog adni a számítógép. A válaszokat természetesen az adatok határozták meg, de azok nem voltak kiszámíthatóak. Kiderült, hogy a *meghatározottság* és a *kiszámíthatóság* közötti különbségtétel fontos, amire még visszatérek.

Emlékszem, milyen izgalmas volt, amikor az előlapi lámpák elsötétültek, közvetlenül azelőtt, hogy az algoritmus befejezte a tanácskozást, mintha a számítógép mélyen elgondolkodott volna. Amikor az emberek odajöttek, és alig várták a következő eredményeket, a finoman villogó fényekre mutattam, és azt mondtam: "Gondolkodik". Ez egyszerre volt és nem is volt vicc - *tényleg* úgy tűnt, hogy a válaszokon gondolkodik -, és a személyzet tagjai elkezdtek személyiséget tulajdonítani a gépnek. Ez talán antropomorfizáció volt, de arra készítetted, hogy komolyan elgondolkodjak a gondolkodás és

a számítástechnika közötti kapcsolaton.

Annak érdekében, hogy felmérjem, hogy a saját agyam mennyire hasonlít az általam ismert számítógépes programokhoz, elkezdtem azon gondolkodni, hogy mit kell tennie az agyumnak az információ feldolgozása során. Ötven éve folytatom ezt a vizsgálatot. Amit az alábbiakban leírok az agy működésének jelenlegi megértéséről, az nagyon másképp hangzik, mint a számítógép szokásos fogalma. Alapvetően azonban az agy valóban információt tárol és dolgoz fel, és a számítás egyetemessége miatt - amelyre szintén vissza fogok térni - az agy és a számítógépek között több a párhuzam, mint amennyire látszik.

Minden alkalommal, amikor teszek valamit - vagy gondolok valamire -, akár fogmosásról, akár a konyhában való sétáról, akár egy üzleti problémán való elmélkedésről, akár a billentyűkön való gyakorlásról, akár egy új ötletről van szó, elgondolkodom azon, hogyan tudtam megvalósítani azt. Még inkább elgondolkodom mindazon, amire nem vagyok képes, mivel az emberi gondolkodás korlátai ugyanolyan fontos támpontokat adnak. Az, hogy ennyit gondolkodom a gondolkodásról, nagyon is lehet, hogy lelassít, de bízom abban, hogy az ilyen önreflexiós gyakorlatok lehetővé teszik számomra, hogy finomítsam a mentális módszereimet.

Ahhoz, hogy tudatosíthassuk magunkban, hogyan működik az agyunk, nézzünk meg egy sor elmekísérletet.

Próbáld ki ezt: *Mondd el az ábécét.*

Valószínűleg emlékszik erre gyermekkorából, és könnyen meg tudja csinálni.

Oké, most próbáld meg ezt: *Mondd el az ábécét visszafelé.*

Hacsak nem ebben a sorrendben tanultad az ábécét, valószínűleg lehetetlennek fogod találni. Alkalmanként előfordul, hogy valaki, aki jelentős időt töltött olyan általános iskolai osztályteremben, ahol az ábécé ki van téve, képes lesz előhívni a vizuális emlékezetét, és onnan visszafelé olvasni. De még ez is nehéz, mert valójában nem emlékszünk egész képekre. Az ábécé visszafelé történő felolvasásának egyszerű feladatnak kellene lennie, hiszen pontosan ugyanazokat az információkat tartalmazza, mint az előre történő felolvasás, mégis általában képtelenek vagyunk rá.

Emlékszik a társadalombiztosítási számára? Ha igen, fel tudja mondani visszafelé anélkül, hogy előbb leírja? Mit szólna a "Marynek volt egy kis báránya" című gyermekdalhoz? A számítógépek triviálisan képesek erre. Mégis kudarcot vallunk, hacsak nem tanuljuk meg kifejezetten a visszafelé történő sorolást, mint új sorozatot. Ez valami fontosat árul el arról, hogy az emberi memória hogyan szerveződik.

Természetesen ezt a feladatot könnyen elvégezhetjük, ha leírjuk a sorrendet, majd visszafelé olvasunk. Ezzel egy technológiát - az írott nyelvet - használunk arra, hogy kompenzáljuk a segédeszköz nélküli gondolkodásunk egyik korlátját, bár ez egy nagyon korai eszköz. (Ez volt a második találmányunk, a beszélt nyelv volt az első.) Ezért találunk fel eszközöket - hogy kompenzáljuk a hiányosságainkat.

Ez azt sugallja, **hogy emlékeink egymás után következnek és sorrendben vannak. Az emlékek abban a sorrendben érhetők el, ahogyan emlékeznek rájuk. Nem tudjuk közvetlenül visszafordítani az emlékek sorrendjét.**

Nekünk is nehézséget okoz egy memória elindítása egy sorozat közepén. Ha

megtanulok zongorán játszani egy zenedarabot, általában nem tudom csak úgy elkezdni a közepén egy tetszőleges ponton. Van néhány pont, ahol be tudok ugrani, mert a darabról való szekvenciális emlékezetem szegmensekbe szerveződik. Ha azonban egy szegmens közepén próbálok belekezdeni, vissza kell térnem a hangjegyolvasáshoz, amíg a szekvenciális memóriám be nem indul.

Ezután próbáld ki ezt: *Emlékezzon vissza egy sétára, amit az elmúlt napokban tett.*

Mire emlékszik róla?

Ez az elmekísérlet akkor működik a legjobban, ha nemrég sétáltál, például ma vagy tegnap (helyettesítheted egy autóvezetéssel, vagy gyakorlatilag bármilyen tevékenységgel, amely során valamilyen terepen mozogtál).

Valószínű, hogy nem sok mindenre emlékszik az élményről. Ki volt az ötödik személy, akivel találkoztál (nem csak az ismerősöket beleértve)? Láttál egy tölgyfát? Egy postaládát? Mit láttál, amikor befordultál az első sarkon? Ha elhaladtál néhány bolt mellett, mi volt a második kirakatban? Talán rekonstruálni tudja a válaszokat néhány ilyen kérdésre abból a néhány nyomból, amelyre emlékszik, de valószínű, hogy viszonylag kevés részletre emlékszik, még akkor is, ha ez egy nagyon friss élmény.

Ha rendszeresen sétál, gondoljon vissza a múlt hónapban tett első sétájára (vagy a múlt hónapban tett első irodai útjára, ha ingázik). Valószínűleg egyáltalán nem emlékszik arra a konkrét sétára vagy ingázásra, és ha mégis, akkor kétségtelenül még kevesebb részletre emlékszik róla, mint a mai sétájáról.

A későbbiekben kitérek a tudatosság kérdésre, és rámutatok arra, hogy hajlamosak vagyunk a tudatosságot az eseményekre vonatkozó emlékezetünkkel egyenlővé tenni. Az elsődleges ok, amiért azt hisszük, hogy nem vagyunk tudatánál, amikor altatásban vagyunk, az az, hogy nem emlékszünk semmire abból az időszakból (bár vannak érdekes - és nyugtalanító - kivételek). Tehát a ma reggeli sétámat illetően, nem voltam tudatánál a nagy részében? Ez egy jogos kérdés, tekintve, hogy szinte semmire sem emlékszem abból, amit láttam, vagy akár arra, hogy mire gondoltam.

Történetesen van néhány dolog, amire emlékszem a ma reggeli sétámról. Emlékszem, hogy gondoltam erre a könyvre, de nem tudnám pontosan megmondani, hogy mik voltak ezek a gondolatok. Arra is emlékszem, hogy elhaladtam egy babakocsit toló nő mellett. Emlékszem, hogy a nő vonzó volt, és hogy a baba is aranyos volt. Két gondolatra emlékszem, ami ezzel az élménnyel kapcsolatban felmerült bennem: *Ez a baba imádnivaló, mint az új unokám, és Mit érzel ez a baba a vizuális környezetében?* Arra nem emlékszem, hogy mit viselt egyikük sem, vagy milyen színű volt a hajuk. (A feleségem majd elmondja, hogy ez jellemző.) Bár képtelen vagyok bármi konkrétumot leírni a külsejükkel kapcsolatban, van valami kimondhatatlan érzésem arról, hogy az anya hogy nézett ki, és azt hiszem, több különböző nő képe közül is ki tudnám választani az ő képét. Tehát bár biztosan van valami a külsejével kapcsolatban, ami megmaradt az emlékezetemben, ha a nőre, a babakocsira és a csecsemőre gondolok, képtelen vagyok elképzelni őket.

Erről az eseményről nincs fénykép vagy videó az emlékezetemben. Nehéz pontosan leírni, hogy mi *van* a fejemben erről az élményről.

Arra is emlékszem, hogy néhány héttel korábban egy másik nő mellett mentem el egy babakocsival sétálva. Abban az esetben nem hiszem, hogy felismerném annak a nőnek a képét. Ez az emlék most sokkal halványabb, mint amilyen röviddel azután a séta után lehetett.

Ezután *gondoljon olyan emberekre, akikkel csak egyszer vagy kétszer találkozott*. Tisztán el tudja őket képzelni? Ha képzőművész vagy, akkor ezt a megfigyelőképességet talán megtanultad, de jellemzően képtelenek vagyunk olyan embereket vizualizálni, akikkel csak véletlenül találkoztunk, hogy lerajzoljuk vagy kellőképpen leírjuk őket, de egy képet felismerni róluk nem okozna nagy nehézséget.

Ez arra utal, hogy az **agyban nincsenek tárolt képek, videók vagy hangfelvételek. Emlékeinket minták sorozataként tároljuk. Az idővel elhalványulnak azok az emlékek, amelyekhez nem férünk hozzá.** Amikor a rendőrségi fantomképrajzolók kihallgatnak egy bűncselekmény áldozatát, nem kérdezik meg, hogy "Hogy nézett ki az elkövető szemöldöke?". Inkább megmutatnak egy sor szemöldökképet, és megkérik az áldozatot, hogy válasszon ki egyet. A megfelelő szemöldöksorozat ugyanannak a mintának a felismerését váltja ki, amely az áldozat emlékezetében tárolódik.

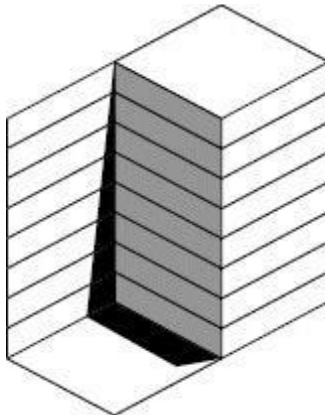
Most nézzük meg azokat az arcokat, amelyeket jól ismersz. *Felismered valamelyiküket?*



Kétségtelenül felismeri ezeket az ismerős személyiségeket, még akkor is, ha részben elfedik vagy eltorzítják őket. Ez az emberi érzékelés egyik legfontosabb erőssége: **Akkor is felismerünk egy mintát, ha annak csak egy részét érzékeljük (látjuk, halljuk, érezzük), és akkor is, ha az elváltozásokat tartalmaz. Felismerő képességünk nyilvánvalóan képes egy minta invariáns jellemzőit felismerni**

-jellemzők, amelyek túlélnek a valós világbeli változásokat. A karikatúrában vagy a művészet bizonyos formáiban, például az impresszionizmusban megjelenő látszólagos torzítások a kép (személy, tárgy) általunk felismert mintáit hangsúlyozzák, miközben más részleteket megváltoztatnak. A művészet világa valójában megelőzte a tudomány világát az emberi érzékelőrendszer erejének megbecsülésében. Ugyanezt a megközelítést alkalmazzuk, amikor néhány hangból felismerünk egy dallamot.

Most nézzük meg ezt a képet:



A kép kétértelmű - a fekete tartomány által jelzett sarok lehet belső sarok vagy külső sarok. Elsőre valószínűleg az egyik vagy a másik irányba érzékeled, bár némi erőfeszítéssel megváltoztathatod az érzékelésedet az alternatív értelmezésre. Ha azonban az elméd egyszer rögzült egy felfogáson, nehéz lehet meglátni a másik perspektívát. (Kiderül, hogy ez igaz az intellektuális perspektívákra is.) A kép agyad általi értelmezése valójában befolyásolja a kép

megtapasztalását. Amikor a sarok belsőnek tűnik, az agyad árnyékként értelmezi a szürke régiót, így nem tűnik olyan sötétnek, mint amikor a sarkot külsőnek értelmezed.

Így az észleléseinkről szerzett tudatos tapasztalatainkat valójában az értelmezéseink változtatják meg.

Vegyük figyelembe, hogy azt látjuk, amire számítunk

Biztos vagyok benne, hogy a fenti mondatot ki tudtad fejezni.

Ha kiírtam volna az utolsó szót, akkor csak egy pillanatra kellett volna rápillantania, hogy megbizonyosodjon arról, hogy ez az, amire számított.

Ez azt jelenti, hogy **folyamatosan jósoljuk a jövőt, és feltételezzük, hogy mit fogunk tapasztalni. Ez a várakozás befolyásolja azt, amit ténylegesen érzékelünk.** A jövő előrejelzése tulajdonképpen az elsődleges oka annak, hogy van agyunk.

Gondoljunk egy olyan tapasztalatra, amelyet mindannyian rendszeresen átélünk: Egy évvel ezelőtti emlék megmagyarázhatatlanul felbukkan a fejünkben.

Ez gyakran egy olyan személy vagy esemény emléke lesz, akire már régóta nem gondolt. Nyilvánvaló, hogy valami kiváltotta az emléket. A gondolatmenet, amely ezt okozta, nyilvánvaló lehet, és olyasmi, amit képes vagy megfogalmazni. Máskor lehet, hogy tudatában van az emlékhöz vezető gondolatsornak, de nehezen tudná kifejezni. Gyakran előfordul, hogy a kiváltó ok gyorsan eltűnik, így az emlék úgy tűnik, mintha a semmiből jött volna. Gyakran tapasztalom ezeket a véletlenszerű emlékeket rutineljárások, például fogmosás közben. Néha tudatában lehetek a kapcsolatnak - a fogkeféről leeső fogkrém emlékeztethet az ecsetről leeső festékre egy főiskolai festőórán, amelyen részt vettem. Néha csak homályosan érzem a kapcsolatot, vagy egyáltalán nem.

Ezzel kapcsolatos jelenség, amelyet mindenki gyakran tapasztal, hogy megpróbál egy nevet vagy egy szót kitalálni. Az eljárás, amit ilyenkor alkalmazunk, az, hogy megpróbáljuk felidézni magunkban azokat a kiváltó okokat, amelyek feloldhatják az emléket. (Például: *Ki játszotta Padmé királynőt A Sith-ek bosszújában? Lássuk csak, ugyanaz a színésznő, aki nemrég egy táncról szóló sötét filmben, a Fekete hattyúban volt a sztár, ó igen, Natalie Portman*). Néha sajátos mnemonikákat veszünk fel, hogy segítsenek emlékezni. (Például: *Igen, Portman, Natalie Portman. Portman.*) Néhányan a emlékeink elég erősek ahhoz, hogy egy kérdéstől (például, *hogy ki játszotta Padmé királynőt*) egyenesen a válaszhoz tudjunk eljutni; gyakran viszont több kiváltó okon kell végigmennünk, amíg megtaláljuk a megfelelőt. Ez nagyon hasonlít a megfelelő weblinkhez. Az emlékek valóban elveszhetnek, mint egy olyan weblap, amelyre nem mutat más oldal (legalábbis nem találunk olyan oldalt, amelyre más oldal mutat).

Rutineljárások végrehajtása közben - például az ing felvétele - figyelje meg, hogyan hajtja végre azokat, és mérlegelje, hogy mennyire követi minden alkalommal ugyanazt a lépéssort. Saját megfigyeléseim alapján (és mint említettem, folyamatosan próbálok megfigyelni magam), valószínű, hogy nagyjából ugyanazokat a lépéseket követi minden egyes alkalommal, amikor egy adott rutinfeladatot végez, bár előfordulhat, hogy további modulok is hozzáadódnak. Például a legtöbb ingemhez nem szükséges mandzsettagomb, de ha mégis, akkor az egy további feladatsorral jár.

A lépések listái a fejemben hierarchiákba rendeződnek. Rutinszerű eljárást követek lefekvés előtt. Az első lépés a fogmosás. De ez a művelet viszont kisebb lépéssorozatra bomlik, amelyek közül az első az, hogy fogkrémet teszek a fogkefére. Ez a lépés viszont még kisebb lépésekből áll, mint például a fogkrém megtalálása, a kupak levétele és így tovább. A fogkrém megtalálásának lépése szintén lépésekből áll, amelyek közül az első a fürdőszobaszekrény kinyitása. Ez a lépés viszont lépéseket igényel, amelyek közül az első a szekrényajtó külső oldalának megfogása. Ez a fészkelődés tulajdonképpen egészen a mozdulatok nagyon finom szemcséiig folytatódik, így szó szerint ezernyi apró művelet alkotja az éjszakai rutinomat. Bár lehet, hogy nehezen emlékszem egy néhány órával ezelőtti séta részleteire, de az ágyba készülődés mindezen sok lépésének felidézése nem okoz nehézséget - olyannyira, hogy képes vagyok más dolgokra is gondolni, miközben ezeket a műveleteket végigcsinálom. Fontos rámutatni, hogy ez a lista nem egyetlen hosszú, több ezer lépésből álló listaként tárolódik el.

-helyesebben, minden egyes rutineljáráskra egymásba ágyazott tevékenységek bonyolult hierarchiájaként emlékszünk.

Ugyanilyen hierarchia érvényesül a tárgyak és helyzetek felismerésének képességében is. Felismerjük az általunk jól ismert emberek arcát, és azt is, hogy ezek az arcok szemeket, orrot, száját és így tovább tartalmaznak - ez egy olyan hierarchikus mintázat, amelyet mindkettőnél használunk.

az észleléseinket és a cselekedeteinket. A hierarchiák használata lehetővé teszi számunkra a minták újrafelhasználását. Például nem kell minden alkalommal, amikor egy új arcot ismerünk meg, újra megtanulnunk az orr és a száj fogalmát.

A [következő fejezetben](#) ezeknek a gondolat kísérleteknek az eredményeit egy elméletbe foglaljuk, amely a neokortex működésére vonatkozik. Amellett fogok érvelni, hogy ezek olyan alapvető tulajdonságokat tárnak fel gondolkodásunkban, amelyek egységesek, a fogkrém megtalálásától a versírásig.

FEJEZET 3

A NEOKORTEX MODELLJE: AZ ELME MINTAFELISMERÉSI ELMÉLETE

Az agy egy szövet. Bonyolult, bonyolult szövésű szövet, mint semmi más, amit az univerzumban ismerünk, de sejtekből áll, mint minden szövet. Ezek ugyan magasan specializált sejtek, de a többi sejtre is érvényes törvények szerint működnek. Elektromos és kémiai jelek érzékelhetők, rögzíthetők és értelmezhetők, vegyi anyagaik azonosíthatók; az agy szövevényes szövedékét alkotó kapcsolatok feltérképezhetők. Röviden, az agyat éppúgy lehet tanulmányozni, mint a vesét.

-David H. Hubel, idegtudós

Tegyük fel, hogy van egy gép, amelynek szerkezete gondolkodást, érzést és érzékelést eredményez; képzeljük el ezt a gépet megnagyobbítva, de ugyanazokat az arányokat megtartva, úgy, hogy úgy léphetnének be, mintha egy malom lenne. Ezt feltételezve, meglátogathatnád a belsejét; de mit észlelhetnél ott? Semmi mást, csak részeket

amelyek egymást lökdörik és mozgatják, és soha semmi olyat, ami megmagyarázná az érzékelést.

-Gottfried Wilhelm Leibniz

A minták hierarchiája

Az [előző fejezetben](#) leírt egyszerű kísérleteket és megfigyeléseket ezerszer megismételtem számtalan kontextusban. E megfigyelésekből levont következtetések szükségszerűen korlátozzák az agy működésére vonatkozó magyarázatomat, ahogyan a XIX. század elején és végén végzett egyszerű kísérletek az idővel, a térrel és a tömeggel kapcsolatban szükségszerűen korlátozták a fiatal Einstein mester gondolatait a világegyetem működéséről. A következő vitában az idegtudomány néhány nagyon alapvető megfigyelését is figyelembe veszem, megkísérelve elkerülni a még mindig vitatott számos részletet.

Először is hadd magyarázzam el, hogy ez a rész miért tárgyalja kifejezetten a neokortexet (a latinból, ami "új kérget" jelent). Tudjuk, hogy a neokortex felelős azért, hogy képesek vagyunk az információmintákat kezelni, mégpedig hierarchikusan. A neokortex nélküli állatok (alapvetően a nem emlősök) nagyrészt képtelenek arra, hogy

hierarchiák megértése.¹ A valóság eredendően hierarchikus természetének megértése és kihasználása egyedülállóan emlősökre jellemző tulajdonság, és annak köszönhető, hogy az emlősök egyedülálló módon rendelkeznek ezzel az evolúciósan új keletű agyi struktúrával. A neokortex felelős az érzékszervi észlelésért, a vizuális tárgytól az absztrakt fogalmakig minden felismeréséért, a mozgás irányításáért, a térbeli tájékozódástól a racionális gondolkodásig terjedő következtetésekért és a nyelvért - lényegében azért, amit mi "gondolkodásnak" tekintünk.

Az emberi neokortex, az agy legkülső rétege vékony, lényegében kétdimenziós struktúra, vastagsága körülbelül 2,5 milliméter (körülbelül egy tized hüvelyk). Rágcsálóknál körülbelül bélyegméretű és sima. A főemlősöknél evolúciós újítás, hogy a felület növelése érdekében bonyolultan hajtogatottá vált az agy többi részének tetején, mély barázdákkal, barázdákkal és ráncokkal. A bonyolult hajtogatásnak köszönhetően a neokortex alkotja az emberi agy nagy részét, súlyának 80 százalékát. A *homo sapiens* nagy homlokot fejlesztett ki, hogy még nagyobb neokortexet tudjon kialakítani; különösen a homloklebenyünk van meg, ahol a magas szintű fogalmakkal kapcsolatos absztraktabb mintákkal foglalkozunk.

Ez a vékony szerkezet alapvetően hat rétegből áll, amelyek I-től (a legkülső réteg) VI-ig terjednek. A II. és III. réteg neuronjaiból kiinduló axonok a neokortex más részeire vetülnek. Az V. és VI. rétegből származó axonok (kimeneti kapcsolatok) elsősorban a neokortexen kívülre, a talamuszhoz, az agytörzshöz és a gerincvelőhöz kapcsolódnak. A IV. réteg neuronjai szinaptikus (bemeneti) kapcsolatokat kapnak a neokortexen kívüli neuronoktól, különösen a talamuszban. A rétegek száma régióként kissé változik. A IV. réteg nagyon vékony a motoros kéregben, mert azon a területen nagyrészt nem kap bemenetet a talamuszból, az agytörzsből vagy a gerincvelőből. Ezzel szemben a nyakszirti lebenyben (a neokortex általában a vizuális feldolgozásért felelős része) a IV. rétegben három további alréteg látható, mivel jelentős bemenet áramlik ebbe a régióba, többek között

a talamuszból.

A neokortexszel kapcsolatos kritikusan fontos megfigyelés az alapvető szerkezetének rendkívüli egységessége. Ezt először Vernon Mountcastle (született 1918-ban) amerikai idegtudós vette észre. Mountcastle 1957 fedezte fel a neokortex oszlopos szerveződését. 1978-ban olyan megfigyelést tett, amely az idegtudományok számára legalább akkora jelentőséggel bír, mint a fizika számára a Michelson-Morley-féle 1887-es éter megcáfoló kísérlet. Abban az évben leírta a neokortex figyelemre méltóan változatlan szerveződését, feltételezve, hogy az egyetlen mechanizmusból áll, amely többször ismétlődik.

és újra és újra,² és az agykérgi oszlopot javasolta alapegységnek. A különböző régiók egyes rétegeinek magasságában a fent említett különbségek egyszerűen a régiók által kezelt összekapcsolhatóság mértékének különbségei.

Mountcastle feltételezte, hogy az oszlopokon belül minioszlopok léteznek, de ez az elmélet vitatottá vált, mivel nem voltak látható elhatárolásai az ilyen kisebb szerkezeteknek. Kiterjedt kísérletek azonban kimutatták, hogy az egyes oszlopok neuronszövetén belül valóban vannak ismétlődő egységek. Állításom szerint az alapegység egy mintafelismerő, és ez alkotja a neokortex alapvető komponensét. Mountcastle minioszlopról alkotott elképzelésével ellentétben ezeknek a felismerőknek nincs konkrét fizikai határa, mivel szorosan egymás mellett helyezkednek el, egymásba fonódva, így az agykérgi oszlop egyszerűen nagyszámú felismerő aggregátuma. Ezek a felismerők az élet folyamán képesek egymáshoz kapcsolódni, így a neokortexben látható bonyolult (modulok közötti) összekapcsoltságot nem a genetikai kód írja elő, hanem inkább úgy jön létre, hogy az idő múlásával ténylegesen megtanult mintákat tükrözze. Ezt a tézist részletesebben is le fogom írni, de fenntartom, hogy a neokortexnek így kell szerveződnie.

Mielőtt tovább vizsgálánánk a neokortex szerkezetét, meg kell jegyeznünk, hogy fontos, hogy a rendszereket a megfelelő szinten modellezzük. Bár a kémia elméletileg a fizikára épül, és teljes egészében levezethető lenne a fizikából, ez a gyakorlatban nehézkes és kivitelezhetetlen lenne, ezért a kémia saját szabályokat és modelleket hozott létre. Hasonlóképpen a termodinamika törvényeit is le kellene tudnunk vezetni a fizikából, de amint elegendő számú részecskével rendelkezünk ahhoz, hogy nem egyszerűen részecskék halmazának, hanem gáznak nevezzük őket, az egyes részecskék kölcsönhatásának fizikájára vonatkozó egyenletek megoldása reménytelenül válik, míg a termodinamika törvényei elég jól működnek. A biológiának is megvannak a maga szabályai és modelljei. Egyetlen hasnyálmirigy-szigetsejt rendkívül bonyolult, különösen, ha a molekulák szintjén modellezzük; annak modellezése, amit a hasnyálmirigy az inzulin és az emésztőenzimek szintjének szabályozása szempontjából ténylegesen csinál, lényegesen kevésbé bonyolult.

Ugyanez az elv érvényes az agyban a modellezés és a megértés szintjeire is. Az agy visszafejtésének kétségtelenül hasznos és szükséges része az agy molekuláris szintű kölcsönhatásainak modellezése, de az erőfeszítések célja itt alapvetően az, hogy finomítsuk a modellünket, hogy számot adjunk arról, hogyan dolgozza fel az agy az információt a kognitív jelentés előállítására érdekében.

Herbert A. Simon (1916-2001) amerikai tudós, akinek a mesterséges intelligencia megalapítójaként tartják számon, ékesszólóan írt a komplex rendszerek megfelelő

absztrakciós szinten történő megértésének kérdéséről. Egy általa kidolgozott, EPAM (elementary perceiver and memorizer) nevű mesterséges intelligenciaprogram leírásakor 1973-ban a következőket írta: "Tegyük fel, hogy úgy döntesz, hogy meg akarod érteni a rejtélyes EPAM programot, ami nekem van. Két változatot is tudnék neked adni belőle. Az egyik lenne... az a forma, amelyben valójában megírták - a rutinok és szubrutinok teljes szerkezetével..... Másik lehetőség, hogy az EPAM egy gépi nyelvű változatát adnám át, miután a teljes fordítást elvégeztem - miután úgymond ellaposítottam..... Azt hiszem, nem kell hosszasan vitatkoznom azon, hogy e két változat közül melyik a legegyszerűbb, a legértelmesebb, a legjogszerűbb leírás.... A harmadikat még csak nem is javaslom..., hogy egyik programot sem adjam meg, hanem helyette az elektromágneses egyenleteket és peremfeltételeket, amelyeknek a számítógépnek, fizikai rendszerként tekintve, engedelmessé kellene, miközben EPAM-ként viselkedik. Ez lenne a redukció csúcspontja és érthetlenség."³

Az emberi neokortexben körülbelül félmillió agykérgi oszlop található, amelyek mindegyike körülbelül két milliméter magas és fél milliméter széles helyet foglal el, és körülbelül 60,000 neuronokat tartalmaz (így összesen körülbelül 30 milliárd neuron található a neokortexben). Durva becslés szerint egy agykérgi oszlopon belül minden egyes mintafelismerő körülbelül 100 neuront tartalmaz, így a neokortexben összesen körülbelül 300 millió mintafelismerő található.

Amikor azt vizsgáljuk, hogyan működnek ezek a mintafelismerők, hadd kezdjem azzal, hogy nehéz pontosan tudni, hol kezdjük. A neokortexben minden egyszerre történik, így nincs kezdete és nincs vége a folyamatoknak. Gyakran kell majd hivatkoznom olyan jelenségekre, amelyeket még nem magyaráztam el, de tervezem, hogy még visszatérek rájuk, ezért kérem, viseljék el ezeket az előremutató hivatkozásokat.

Az emberi lények csak gyenge logikai feldolgozó képességgel rendelkeznek, de nagyon mélyen rejlik bennük a minták felismerésének képessége. A logikus gondolkodáshoz a neokortexet kell használnunk, amely alapvetően egy nagy mintafelismerő. Ez nem ideális mechanizmus a logikai átalakítások elvégzésére, de ez az egyetlen eszközünk erre a feladatra. Hasonlítsuk össze például, hogyan sakkozik egy ember, és hogyan működik egy tipikus számítógépes sakkprogram. A Deep Blue, a számítógép, amely 1997-ben legyőzte Garri Kaszparovot, az emberi sakkvilágbajnokot, képes volt másodpercenként 200 millió táblaállás (különböző lépés-ellenlépés szekvenciák) logikai következményeinek elemzésére. (Ez egyébként ma már néhány személyi számítógépen is megoldható.) Kaszparovot megkérdezték, hogy hány állást tudott másodpercenként elemezni, és azt mondta, hogy kevesebb mint egyet. Hogyan lehetséges akkor, hogy egyáltalán képes volt tartani a Deep Blue-val szemben? A válasz az a nagyon erős képesség, amellyel az emberek rendelkeznek a minták felismerésére. Ezt a képességet azonban gyakorolni kell, ezért nem mindenki tud mestersakkozni.

Kaszparov körülbelül 100 000 táblaállást tanult meg. Ez egy valós szám - megállapítottuk, hogy egy emberi mester egy adott területen körülbelül 100 000 tudásdarabot sajátított el. Shakespeare a darabjait 100,000 szóérzékkal komponálta

(körülbelül 29 000 különböző szót használva, de ezek többségét többféleképpen is). Az emberi orvosok tudásának reprezentálására épített orvosi szakértői rendszerek azt mutatták, hogy egy tipikus emberi orvossal körülbelül 100 000 fogalmat sajátított el a saját területén. Egy tudásdarab felismerése ebből a tárházból nem egyszerű, mivel egy adott elem minden egyes alkalommal egy kicsit másképp jelenik meg, amikor megtapasztaljuk.

Tudásával felvértezve Kaszparov ránéz a sakk táblára, és az általa látott mintákat összehasonlítja mind a 100 000 táblás szituációval, amelyeket elsajátított, és az összes összehasonlítást 100,000 egyszerre végzi. Ebben a kérdésben egyetértés van: Minden neuronunk egyszerre dolgozza fel - veszi figyelembe a mintákat - a mintákat. Ez nem azt jelenti, hogy mind egyszerre *tüzelnek* (valószínűleg a padlóra esnének, ha ez megtörténne), de a feldolgozásuk közben mérleget a tüzelés lehetőségét.

Hány mintát képes tárolni a neokortex? Figyelembe kell vennünk a redundancia jelenségét. Egy szeretett személy arcát például nem egyszer, hanem több ezerszer tároljuk el. Ezen ismétlések némelyike nagyrészt ugyanazt az arcképet mutatja, míg a legtöbb más-más perspektívából, más megvilágításban, más arckifejezéssel stb. mutatja az arcot. Ezek közül az ismétlődő minták közül egyiket sem tároljuk önmagában képként (azaz képpontok kétdimenziós tömbjeként). Inkább jellemzők listájaként tárolódnak, ahol a minták alkotóelemei maguk is minták. Az alábbiakban pontosabban leírjuk, hogyan néznek ki és hogyan szerveződnek a jellemzők e hierarchiái.

Ha úgy vesszük, hogy egy szakértő alapvető tudása körülbelül 100 000 "tudásdarabból" (azaz mintából) áll, és a redundancia becsült értéke körülbelül 100, akkor ez 10 millió mintát jelent. Ez az alapvető szakértői tudás általánosabb és szélesebb körű szakmai tudásra épül, így a minták nagyságrendjét körülbelül 30-50 millióra növelhetjük. A hétköznapi "józan emberi" tudásunk még ennél is nagyobb; az "utcai okosság" valójában lényegesen több neokortexet igényel, mint a "könyves okosság". Ha ezt is figyelembe vesszük, akkor becslésünk jóval 100 millió fölé emelkedik, figyelembe véve a kb. 100 millió minta redundanciáját. 100.

-A nagyon gyakori minták redundanciája több ezerre tehető, míg egy vadonatúj jelenség redundanciája kevesebb mint 10 lehet.

Amint azt alább kifejtem, eljárásaink és cselekvéseink is mintákból állnak, és ugyanígy az agykéreg régióiban tárolódnak, így becslésem szerint az emberi neokortex teljes kapacitása több százmillió minta alacsony százmillió nagyságrendű. Ez a durva számadat jól korrelál a mintafelismerők számával, amelyet fentebb körülbelül 300 millióra becsültem, így ésszerű következtetés, hogy minden egyes neokortexi mintafelismerő funkciója egy minta egy ismétlésének (azaz a legtöbb minta többszörösen redundáns példánya közül egy példánynak a neokortexben) a feldolgozása. Az emberi agy által feldolgozni képes minták számára (beleértve a szükséges redundanciát is) és a fizikai mintafelismerők számára vonatkozó becsléseink történetesen azonos nagyságrendűek. Itt kell megjegyezni, hogy amikor egy minta "feldolgozására" utalok, akkor mindarra a dologra gondolok, amit egy mintával képesek vagyunk tenni: megtanulni, megjósolni (beleértve annak részeit is), felismerni és megvalósítani (akár további gondolkodással, akár egy fizikai mozgásmintán keresztül).

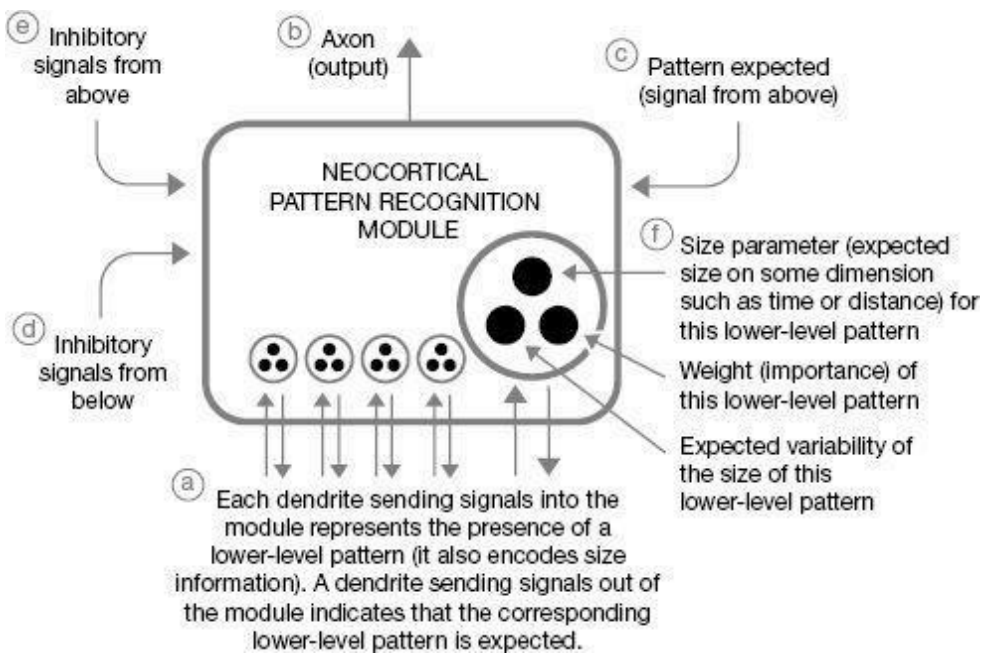
Háromszázmillió mintafeldolgozó nagy számnak tűnhet, és valóban elegendő volt

ahhoz, hogy a *Homo sapiens* kifejllessze a szóbeli és írásbeli nyelvet, az összes eszközünket és más különféle alkotásokat. Ezek a találmányok egymásra épültek, ami a technológiák információtartalmának exponenciális növekedését eredményezte, ahogyan azt a gyorsuló megtérülés törvényében leírtam. Ezt egyetlen más faj sem érte el. Amint azt tárgyaltam, néhány más faj, például a csimpánzok, úgy tűnik, kezdetlegesen képesek a nyelv megértésére és kialakítására, valamint kezdetleges eszközök használatára. Végül is nekik is van neokortexük, de képességeik korlátozottak a kisebb méretük miatt, különösen a homloklebenyben. Saját neokortexünk mérete túllépett egy olyan küszöböt, amely lehetővé tette fajunk számára, hogy egyre erősebb eszközöket építsen, köztük olyanokat is, amelyekkel ma már képesek vagyunk megérteni saját intelligenciánkat. Végső soron az agyunk, az általa elősegített technológiákkal kombinálva, lehetővé fogja tenni számunkra, hogy létrehozzunk egy szintetikus neokortexet, amely jóval több mint 300 millió mintafeldolgozót tartalmaz majd. Miért ne lehetne egymilliárd? Vagy egy trillió?

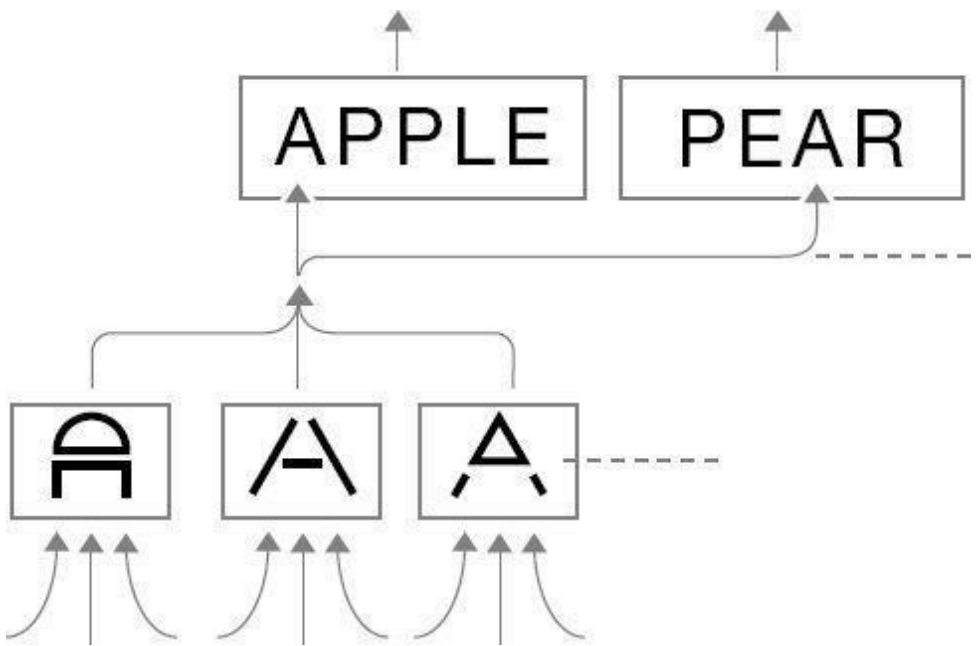
A minta szerkezete

Az általam itt bemutatott mintafelismerési elmélet a neokortex mintafelismerő moduljai által végzett mintafelismerésen alapul. Ezek a minták (és a modulok) hierarchiákba szerveződnek. Az alábbiakban ennek az elképzelésnek a szellemi gyökereit tárgyalom, beleértve a hierarchikus mintafelismeréssel kapcsolatos saját munkámat az 1980-as és 1990-es években, valamint Jeff Hawkins (1957-ben született) és Dileep George (1977-ben született) neokortex-modelljét a 2000-es évek elején.

Minden egyes minta (amelyet a neokortexben található, becslések szerint 300 millió mintafelismerő egyike ismer fel) három részből áll. Az első rész a bemenet, amely a fő mintát alkotó alacsonyabb szintű mintákból áll. Az egyes alacsonyabb szintű minták leírását nem kell megismételni minden egyes magasabb szintű minta esetében, amely hivatkozik rájuk. Például a szavakhoz tartozó minták közül sokban szerepel az "A" betű. Ezen minták mindegyikének nem kell megismételnie az "A" betű leírását, de ugyanazt a leírást kell használniuk. Gondoljon erre úgy, mint egy webmutatóra. Az "A" betűnek van egy weboldala (azaz egy mintája), és az "A" betűt tartalmazó szavak összes weboldala (mintája) hivatkozik az "A" oldalra (az "A" mintára). A neokortex a webes hivatkozások helyett tényleges idegi kapcsolatokat használ. Az "A" mintafelismerőből egy axon indul, amely több dendrithez kapcsolódik, minden egyes "A"-t tartalmazó szóhoz egy-egyhez. Tartsuk szem előtt a redundancia tényezőt is: Több mint egy mintafelismerő van az "A" betűhöz. E több "A" mintafelismerő bármelyike küldhet jelet az "A" betűt tartalmazó mintafelismerőkhöz.



Az egyes minták második része a minta neve. A nyelv világában ez a magasabb szintű minta egyszerűen az "alma" szó. Bár közvetlenül a neokortexünket használjuk a nyelv minden szintjének megértésére és feldolgozására, a benne található minták többsége önmagában nem nyelvi minta. A neokortexben egy minta "neve" egyszerűen az az axon, amely az egyes mintafeldolgozókból kilép; amikor ez az axon tüzel, a megfelelő mintát felismertük. Az axon tüzelése az a mintafelismerő, amelyik a minta nevét kiáltja: "Hé, srácok, most láttam meg az "alma" írott szót."



Az "A" három redundáns (de némileg eltérő) mintája, amelyek az "A"-t tartalmazó magasabb szintű mintákat táplálják.

Az egyes minták harmadik és egyben utolsó része a magasabb szintű minták halmaza, amelyeknek viszont részei. Az "A" betű esetében ez az összes olyan szó, amely tartalmazza az "A"-t. Ezek megint csak olyanok, mint a webes hivatkozások. Minden egyes felismert minta egy szinten kiváltja a következő szinten, hogy az adott magasabb szintű minta egy része jelen van. A neokortexben ezeket a linkeket fizikai dendritek reprezentálják, amelyek az egyes agykérgi mintafelismerő neuronokba áramlanak. Ne feledjük, hogy minden neuron több dendritről is kaphat bemenetet, mégis egyetlen kimenetet állít elő egy axonon. Ez az axon viszont több dendritre is továbbíthat.

Hogy néhány egyszerű példát említsünk, a következő oldalon látható egyszerű minták a nyomtatott betűkhöz használt minták egy kis részhalmaza. Vegyük észre, hogy minden szint egy mintát alkot. Ebben az esetben az alakzatok minták, a betűk minták, és a szavak is minták. Mindegyik mintának van egy sor bemenete, egy mintafelismerési folyamata (a bemenetek alapján, amelyek a modulban zajlanak), és egy kimenete (amely a mintafelismerő következő, magasabb szintjére táplálja).

Délnyugati és észak-középső kapcsolat:

Délkeleti és észak-középső kapcsolat:

Vízszintes keresztléc:

A bal szélső függőleges vonal:

Dél felé néző homorú terület:

Alsó vízszintes vonal:

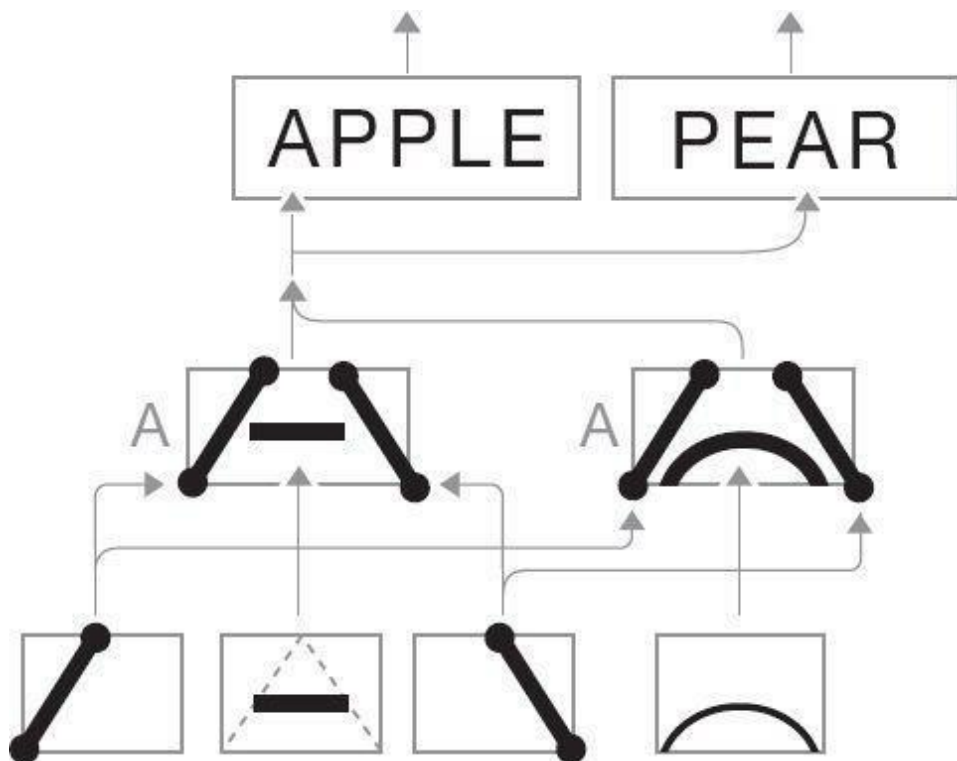
Felső vízszintes vonal:

Középső vízszintes vonal:

A felső régiót alkotó hurok:

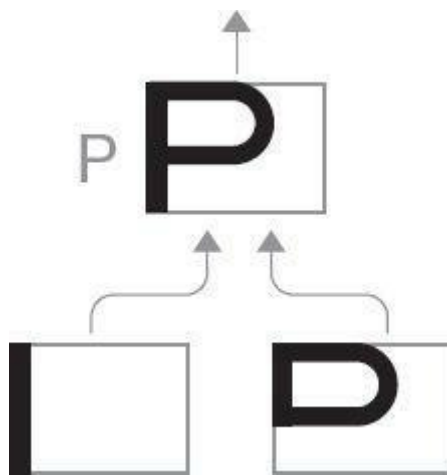
A fenti minták a következő, magasabb szintű minta alkotóelemei, amely a nyomtatott betűk kategóriája (a neokortexen belül azonban nincs ilyen formális kategória; valójában nincsenek is formális kategóriák).

"A":



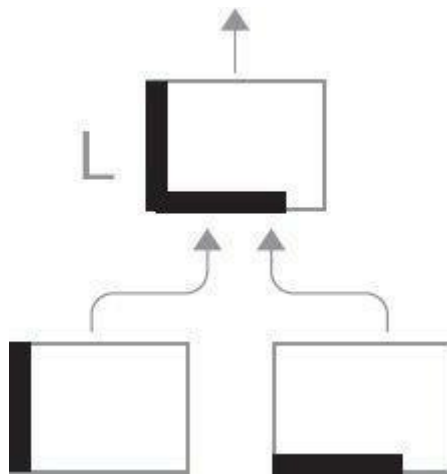
Két különböző minta, amelyek közül bármelyik alkotja az "A"-t, és két különböző magasabb szintű minta ("APPLE" és "PEAR"), amelyeknek az "A" része.

"P":



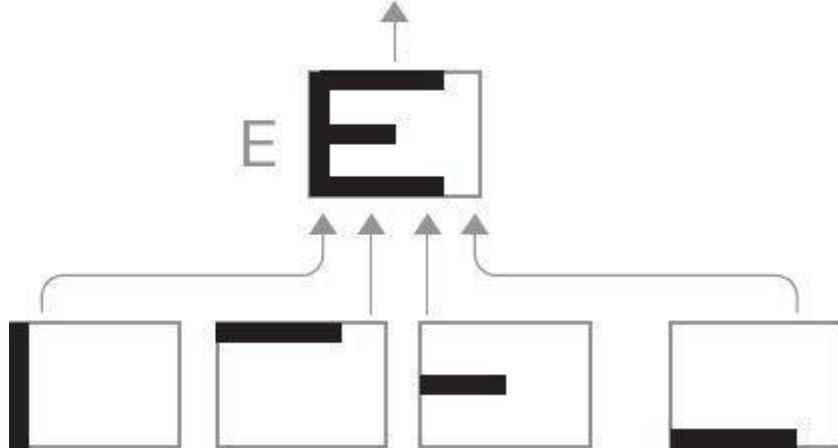
A magasabb szintű "P" minta részét képező minták.

"L":



A magasabb szintű "L" minta részét képező minták.

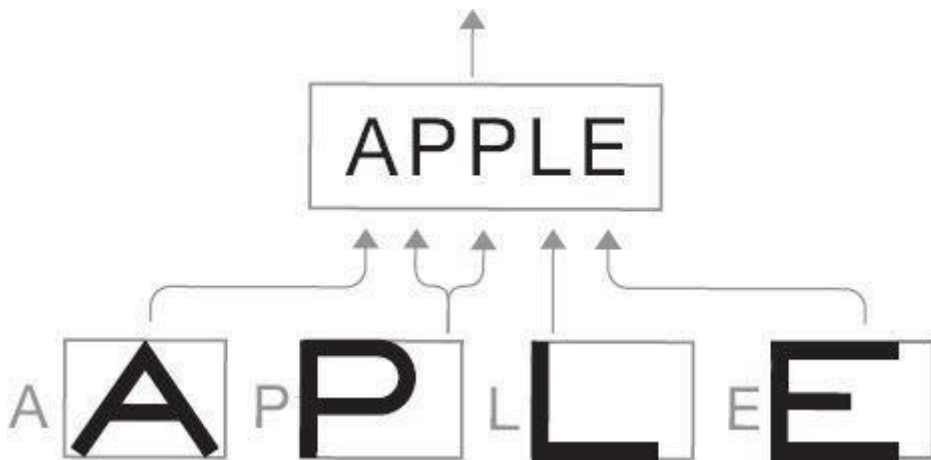
"E":



Az "E" magasabb szintű minta részét képező minták.

Ezek a betűminták egy még magasabb szintű mintázatba, a szavaknak nevezett kategóriába táplálkoznak. (A "szavak" szó a mi nyelvi kategóriánk erre a fogalomra, de a neokortex csak mintaként kezeli őket).

"APPLE":



Az agykéreg egy másik részén található a tárgyak tényleges képeit (szemben a nyomtatott betűkkel) feldolgozó mintafelismerők hasonló hierarchiája. Ha egy valódi almát nézünk, az alacsony szintű felismerők érzékelik az ívelt éleket és a felszíni színmintákat, ami odáig vezet, hogy egy mintafelismerő kilövi az axonját, és azt mondja: "Hé, srákok, most láttam egy valódi almát". Más mintafelismerők viszont hangfrekvenciák kombinációit érzékelik, amelyek a hallókéregben lévő mintafelismerőhöz vezetnek, amely axonjait tüzelve

azt jelezheti: "Most hallottam az "alma" szót."

Ne feledje a redundancia tényezőt - nem csak egyetlen mintafelismerő programunk van az "alma" minden formájára (írott, beszélt, vizuális). Valószínűleg több száz ilyen felismerő tüzel, ha nem több. A redundancia nemcsak annak a valószínűségét növeli, hogy sikeresen felismeri az alma minden egyes példányát, hanem a valós almák változatosságával is foglalkozik. Az almaobjektumok esetében lesznek olyan mintafelismerők, amelyek az alma sokféle formájával foglalkoznak: különböző nézetek, színek, árnyalatok, formák és fajták.

Ne feledje azt sem, hogy a fenti hierarchia a *fogalmak* hierarchiája. Ezek a felismerők fizikailag nem egymás fölött helyezkednek el; a neokortex vékony felépítése miatt fizikailag csak egy mintafelismerővel magasabban van. A fogalmi hierarchiát az egyes mintafelismerők közötti kapcsolatok hozzák létre.

A PRTM fontos jellemzője, hogy a felismerések hogyan történnek az egyes mintafelismerő modulokon belül. A modulban minden egyes bemeneti dendrithez egy súlyt tárolnak, amely jelzi, hogy az adott bemenet mennyire fontos a felismerés szempontjából. A mintafelismerőnek van egy küszöbértéke a tüzeléshez (ami azt jelzi, hogy az adott mintafelismerő sikeresen felismerte az adott mintát, amiért felelős). Nem minden bemeneti mintának kell jelen lennie ahhoz, hogy a felismerő tüzeljen. A felismerő akkor is tüzelhet, ha egy kis súlyú bemenet hiányzik, de kevésbé valószínű, hogy tüzel, ha egy nagy jelentőségű bemenet hiányzik. Amikor tüzel, a mintafelismerő alapvetően azt mondja: "A minta, amelyért felelős vagyok, valószínűleg jelen van".

Egy modul sikeres felismerése a saját mintázatának túlmutat az aktivált bemeneti jelek számolásán (akár a fontossági paraméterrel súlyozott számoláson is). A méret (minden egyes bemenet) számít. Van egy másik paraméter (minden egyes bemenethez), amely a bemenet várható méretét jelzi, és még egy másik, amely azt jelzi, hogy ez a méret mennyire változó. Hogy érzékeltessük, hogyan működik ez, tegyük fel, hogy van egy mintafelismerőnk, amely a "meredek" elhangzott szó felismeréséért felelős. Ez a kimondott szó négy hangból áll: [s], [t], [E] és [p]. A [t] fonéma az úgynevezett "dentális mássalhangzó", ami azt jelenti, hogy a nyelv egy hangkitörést hoz létre, amikor a levegő megszakítja a felső fogakkal való érintkezését. A [t] fonémát lényegében lehetetlen lassan artikulálni. A [p] fonéma "ploszív mássalhangzónak" vagy "szájzárónak" tekinthető, ami azt jelenti, hogy akkor jön létre, amikor a hangszalag hirtelen elzáródik (a [p] esetében az ajkak által), így a levegő már nem jut át rajta. Ez is szükségszerűen gyors. Az [E] magánhangzót a hangszalag és a nyitott száj rezonanciája okozza. "Hosszú magánhangzónak" számít, ami azt jelenti, hogy sokkal hosszabb ideig áll fenn, mint az olyan mássalhangzók, mint a [t] és a [p]; időtartama azonban igen változó lehet. Az [s] fonéma "sziszegő mássalhangzó" néven ismert, és a levegőnek a szorosán egymás mellett tartott fogak széleihez való áramlása okozza. Az időtartama általában rövidebb, mint egy hosszú magánhangzóé, például az [E]-é, de ez is változó (más szóval az [s]-t lehet gyorsan kimondani, de lehet elhúzni is).

A beszédfelismeréssel kapcsolatos munkánk során megállapítottuk, hogy a beszédminták felismeréséhez szükség van az ilyen típusú információk kódolására. Például a "step" és a "steep" szavak nagyon hasonlóak. Bár a "step" [e] fonéma a "step"-ben és az [E] a "steep"-ben némileg eltérő magánhangzók (mivel eltérő rezonanciafrekvenciával

rendelkeznek), mégsem megbízható e két szó megkülönböztetése e gyakran összetéveszthető magánhangzók alapján. Sokkal megbízhatóbb az a megfigyelés, hogy a "step"-ben lévő [e] viszonylag rövid a "steep"-ben lévő [E]-hez képest.

Az ilyen típusú információt minden egyes bemenethez két számmal kódolhatjuk: a várható méret és a méret változékonyságának mértéke. A mi "meredek" példánkban a [t] és a [p] mindkettőnek nagyon rövid várható időtartama, valamint kis várható változékonysága lenne (vagyis nem várjuk, hogy hosszú t-eket és p-eket halljunk). Az [s] hangnak rövid lenne a várható időtartama, de nagyobb a variabilitása, mivel lehetséges, hogy elhúzódik. Az [E] hangnak hosszú várható időtartama, valamint nagyfokú variabilitása van.

Beszédpéldáinkban a "méret" paraméter az időtartamra utal, de az idő csak egy lehetséges dimenzió. A karakterfelismeréssel kapcsolatos munkánk során azt találtuk, hogy a nyomtatott betűk felismeréséhez fontosak az összehasonlítható térbeli információk (például az "i" betű feletti pont várhatóan sokkal kisebb lesz, mint a pont alatti rész). Az absztrakció sokkal magasabb szintjein a neokortex mindenféle folytonosságú mintákkal foglalkozik, például a vonzerő, az irónia, a boldogság, a frusztráció és számtalan más szintjével. Meglehetősen változatos kontinuumok között vonhatunk le hasonlóságokat, ahogy Darwin is tette, amikor a geológiai kanyonok fizikai méretét a fajok közötti differenciálódás mértékével hozta összefüggésbe.

Egy biológiai agyban ezeknek a paramétereknek a forrása az agy saját tapasztalataiból származik. Nem születünk a fonémák veleszületett ismeretével; sőt, a különböző nyelvek nagyon különböző készletekkel rendelkeznek. Ez azt jelenti, hogy egy minta több példája kódolva van az egyes mintafelismerők megtanult paramétereiben (mivel egy minta több példányára van szükség ahhoz, hogy megállapítsuk a minta bemeneti értékeinek várható nagyságeloszlását). Egyes mesterséges intelligencia rendszerekben az ilyen típusú paramétereket szakértők kódolják kézzel (például nyelvészek, akik meg tudják mondani a különböző fonémák várható időtartamát, ahogy azt fentebb kifejtettem). A saját munkám során azt találtuk, hogy az, hogy egy mesterséges intelligencia-rendszer saját maga fedezi fel ezeket a paramétereket a képzési adatokból (hasonlóan ahhoz, ahogyan az agy teszi), jobb megközelítés. Néha hibrid megközelítést alkalmaztunk, azaz a rendszert emberi szakértők intuíciójával alapoztuk meg (a paraméterek kezdeti beállításaihoz), majd a mesterséges intelligencia rendszer automatikusan finomította ezeket a becsléseket valós beszédpéldákból történő tanulási folyamat segítségével.

A mintafelismerő modul kiszámítja annak a valószínűségét (azaz az összes korábbi tapasztalatán alapuló valószínűségét), hogy a minta, amelynek felismeréséért felelős, jelenleg valóban az aktív bemenetei által reprezentált minta. A modul minden egyes bemenete akkor aktív, ha a megfelelő alacsonyabb szintű mintafelismerő tüzel (ami azt jelenti, hogy az adott alacsonyabb szintű mintát felismerték). Minden bemenet kódolja a megfigyelt méretet is (valamilyen megfelelő dimenzióban, mint például az időbeli időtartam vagy a fizikai nagyság vagy valamilyen más kontinuum), hogy a modul összehasonlíthassa a méretet (az egyes bemenetekhez tárolt méretparaméterekkel) a minta általános valószínűségének kiszámításakor.

Hogyan számítja ki az agy (és hogyan tudja egy mesterséges intelligencia rendszer) annak általános valószínűségét, hogy a minta (amelynek felismeréséért a modul felelős)

jelen van, figyelembe véve (1) a bemeneteket (mindegyiknek van egy megfigyelt mérete), (2) a méretre vonatkozó tárolt paramétereket (a várható méret és a méret változékonysága) minden bemenetre, és (3) az egyes bemenetek fontosságának paramétereit? Az 1980-as és 1990-es években én és mások úttörő munkát végeztünk egy hierarchikus rejtett Markov-modelleknek nevezett matematikai módszerrel, amellyel megtanulhatjuk ezeket a paramétereket, majd felhasználhatjuk őket hierarchikus minták felismerésére. Ezt a technikát az emberi beszéd felismerésében, valamint a természetes nyelv megértésében használtuk. Ezt a megközelítést a [7. fejezetben](#) ismertetem részletesebben.

Visszatérve a felismerés áramlására a mintafelismerők egyik szintjéről a másikra, a fenti példában az információáramlás a fogalmi hierarchiában az alapvető betűjellemezőktől a betűkig és a szavakig tart. A felismerések onnan tovább áramlanak felfelé a mondatokig, majd az összetettebb nyelvi szerkezetekig. Ha még több tucat szintet megyünk felfelé, akkor olyan magasabb szintű fogalmakhoz jutunk, mint az irónia és az irigység. Bár minden mintafelismerő egyszerre dolgozik, időbe telik, amíg a felismerések felfelé haladnak ebben a fogalmi hierarchiában. Az egyes szinteken való átjutás néhány század és néhány tizedmásodperc közötti időt vesz igénybe. Kísérletek kimutatták, hogy egy közepesen magas szintű minta, például egy arc legalább egy tizedmásodpercig tart. Jelentős torzulások esetén akár egy teljes másodpercig is eltarthat. Ha az agy szekvenciális lenne (mint a hagyományos számítógépek), és az egyes mintafelismeréseket egymás után végeznék, akkor minden lehetséges alacsony szintű mintát figyelembe kellene vennie, mielőtt a következő szintre lépne. Így sok millió ciklusra lenne szükség, hogy minden egyes szinten végigmenjen. Pontosan ez történik, amikor ezeket a folyamatokat számítógépen szimuláljuk. Ne feledjük azonban, hogy a számítógépek több milliószor gyorsabban dolgoznak, mint a biológiai áramköreink.

Nagyon fontos megjegyezni, hogy az információ a fogalmi hierarchiában lefelé és felfelé is áramlik. Ha valami, akkor ez a lefelé irányuló áramlás még jelentősebb. Ha például balról jobbra haladva olvasunk, és már láttuk és felismertük az "A", "P", "P" és "L" betűket, akkor az "APPLE" felismerő megjósolja, hogy a következő pozícióban valószínűleg egy "E"-t fog látni. Egy jelet küld *le* az "E" felismerőnek, amely gyakorlatilag azt mondja: "Kérem, vegye figyelembe, hogy nagy valószínűséggel hamarosan látni fogja az "E" mintát, ezért legyen résen". Az "E" felismerő ezután úgy állítja be a küszöbértékét, hogy nagyobb valószínűséggel ismerje fel az "E"-t. Ha tehát ezután egy olyan kép jelenik meg, amely homályosan hasonlít egy "E"-re, de talán olyan elmosódott, hogy nem ismerte volna fel "E"-nek...

"normális" körülmények között az "E" felismerő mégis jelezheti, hogy valóban látott egy "E"-t, mivel az várható volt.

A neokortex tehát előre jelzi, hogy mire számít. A jövő előrejelzése az egyik fő oka annak, hogy neokortexünk van. A legmagasabb fogalmi szinten folyamatosan jóslatokat készítünk - ki fog legközelebb belépni az ajtón, mit fog valaki valószínűleg legközelebb mondani, mit várunk, hogy mit fogunk látni, amikor befordulunk a sarkon, a saját cselekedeteink valószínű eredményeit, és így tovább. Ezek a jóslatok folyamatosan zajlanak a neokortex hierarchiájának *minden* szintjén. Gyakran azért ismerünk fel rosszul embereket, dolgokat és szavakat, mert túl alacsony a küszöbünk egy várt minta megerősítésére.

A pozitív jelek mellett vannak negatív vagy gátló jelek is, amelyek azt jelzik, hogy egy bizonyos minta létezése kevésbé valószínű. Ezek származhatnak alacsonyabb fogalmi szintekről (például a bajusz felismerése gátolja annak valószínűségét, hogy a pénztárban álló személy, akit látok, a feleségem), vagy magasabb szintről (például tudom, hogy a feleségem elutazik, tehát a pénztárban álló személy nem lehet ő). Amikor a mintafelismerő gátló jelet kap, megemeli a felismerési küszöböt, de még mindig lehetséges, hogy a minta elsül (tehát ha a sorban álló személy valóban ő, akkor is felismerhetem őt).

A Természet of a Adatok Folyamatos a egy neokortikális mintafelismerőbe

Nézzük tovább, hogyan néznek ki egy minta adatai. Ha a minta egy arc, akkor az adatok legalább két dimenzióban léteznek. Nem mondhatjuk, hogy a szemek feltétlenül az elsők, majd az orr, és így tovább. Ugyanez igaz a legtöbb hangra is. Egy zeneműnek legalább két dimenziója van. Lehet, hogy egynél több hangszer és/vagy hang szólal meg egyszerre. Ráadásul egy összetett hangszer, például a zongora egyetlen hangja több frekvenciából áll. Egyetlen emberi hang egyszerre több tucat különböző frekvenciasávban különböző energiaszintekből áll. Tehát egy hangminta bármelyik pillanatban összetett lehet, és ezek az összetett pillanatok időben elnyúlnak. A tapintási bemenetek szintén kétdimenziósak, mivel a bőr kétdimenziós érzékszerv, és az ilyen minták az idő harmadik dimenziójában változhatnak.

Úgy tűnik tehát, hogy a neokortex mintafeldolgozójának bemenete két-, ha nem háromdimenziós mintákból kell, hogy álljon. A neokortex szerkezetében azonban azt látjuk, hogy a mintázatok bemenetei csak egydimenziós listák. A mesterséges mintafelismerő rendszerek (például beszédfelismerő és vizuális felismerő rendszerek) létrehozása terén végzett összes munkánk azt mutatja, hogy két- és háromdimenziós jelenségeket ilyen egydimenziós listákkal is lehet (és lehetett) reprezentálni. A [fejezetben7](#) leírom, hogyan működnek ezek a módszerek, de egyelőre azzal a felfogással haladhatunk, hogy az egyes mintafeldolgozók bemenete egy egydimenziós lista, még akkor is, ha maga a minta természeténél fogva egynél több dimenziót tükrözhet.

Ezen a ponton figyelembe kell vennünk azt a felismerést, hogy a minták, amelyeket megtanultunk felismerni (például egy konkrét kutya vagy a "kutya" általános fogalma, egy hangjegy vagy egy zenemű), pontosan ugyanaz a mechanizmus, amely az emlékeink alapját

képezi. Emlékeink valójában listákba rendezett minták (ahol minden egyes lista minden egyes eleme egy másik minta az agykérgi hierarchiában), amelyeket megtanultunk, majd felismerünk, amikor megfelelő ingerrel találkozunk. Valójában az emlékek azért léteznek a neokortexben, hogy felismerhetők legyenek.

Ez alól az egyetlen kivétel a lehető legalacsonyabb fogalmi szinten van, ahol a minta bemeneti adatai konkrét érzékszervi információkat jelentenek (például a látóidegből származó képi adatokat). Azonban még ez a legalacsonyabb szintű minta is jelentősen átalakul egyszerű mintázattá, mire az eléri az agykérget. Az emlékezetet alkotó mintázatok listái előremenő sorrendben vannak, és csak ebben a sorrendben vagyunk képesek emlékezni az emlékeinkre, innen ered a nehézségeink az emlékeink visszafordításában.

Egy emléket egy másik gondolatnak/emléknek kell kiváltania (ezek ugyanazok). Ezt a kiváltó mechanizmust akkor tapasztalhatjuk, amikor egy mintát érzékelünk. Amikor érzékeltük az "A", "P", "P" és "L" mintát, az "A P P P L L E" minta megjósolta, hogy egy "E"-t fogunk látni, és kiváltotta az "E" mintát, hogy ez most várható. Az agykéregünk tehát "gondolkodik" azon, hogy egy "E"-t látunk, még mielőtt látnánk. Ha ez a bizonyos interakció az agykéregben a figyelmünket magára vonja, akkor már az "E"-re fogunk gondolni, mielőtt meglátnánk, vagy még akkor is, ha soha nem látjuk. Hasonló mechanizmus váltja ki a régi emlékeket. Általában ilyen kapcsolatok egész láncolata van. Még ha van is valamilyen szintű tudatosságunk a régi emléket kiváltó emlékekről (vagyis a mintákról), az emlékeknek (mintáknak) nincs nyelvi vagy képi címkézésük. Ez az oka annak, hogy a régi emlékek látszólag hirtelen ugranak be a tudatosságunkba. Mivel eltemetve voltak, és talán évekig nem aktiválódtak, ugyanúgy szükségük van egy kiváltó okra, mint ahogy egy weboldalnak szüksége van egy weblinke ahhoz, hogy aktiválódjon. És ahogyan egy weboldal is "elárvulhat", mert nincs más oldal, amelyik hivatkozik rá, ugyanez történhet az emlékeinkkel is.

Gondolataink nagyrészt kétféleképpen aktiválódnak: irányítatlanul és irányítottan, és mindkettő ugyanazokat az agykérgi kapcsolatokat használja. A nem irányított módban hagyjuk, hogy a kapcsolatok maguktól játszódjanak le anélkül, hogy megpróbálnánk őket valamilyen meghatározott irányba mozgatni. A meditáció egyes formái (mint például a Transzcendentális Meditáció, amit én is gyakorlok) pontosan ezen alapulnak, hogy hagyjuk az elmét ezt tenni. Az álmok is rendelkeznek ezzel a tulajdonsággal.

Az irányított gondolkodás során megpróbálunk egy rendezettebb folyamaton keresztül haladni egy emlék (például egy történet) felidézése vagy egy probléma megoldása során. Ez szintén a neokortexünkben lévő listák végiglépegetésével jár, de az irányítatlan gondolkodás kevésbé strukturált pörgése is végigkíséri a folyamatot. Gondolkodásunk teljes tartalma tehát nagyon rendezetlen, ezt a jelenséget James Joyce világította meg "tudatfolyam" című regényeiben.

Ahogy végiggondolod az életedben lévő emlékeket/történeteket/mintákat, legyen szó akár egy véletlen találkozásról egy babakocsis anyával és egy sétáló kisbabával, vagy arról a fontosabb elbeszélésről, hogy hogyan találkoztál a házastársaddal, az emlékeid minták

sorozatából állnak. Mivel ezek a minták nincsenek szavakkal, hangokkal, képekkel vagy videókkal felcímkézve, amikor megpróbálsz felidézni egy jelentős eseményt, lényegében a képeket rekonstruálsz az elmédben, mert a tényleges képek nem léteznek.

Ha "olvasni" akarnánk valakinek az elméjében, és pontosan megnéznénk, hogy mi zajlik a neokortexében, nagyon nehéz lenne értelmezni az emlékeit, akár olyan mintákat néznénk, amelyek egyszerűen csak tárolódnak a neokortexben, és arra várnak, hogy elinduljanak, akár olyanokat, amelyek már elindultak, és jelenleg aktív gondolatokként élnek meg. Amit "látnánk", az a mintafelismerők millióinak egyidejű aktiválódása. Egy századmásodperccel később már egy másik, hasonló számú aktivált mintafelismerőt látnánk. Minden ilyen minta más minták listája lenne, és ezek mindegyike más minták listája lenne, és így tovább, amíg el nem érjük a legelemibb egyszerű mintákat a legalacsonyabb szinten. Rendkívül nehéz lenne értelmezni, hogy mit jelentenek ezek a magasabb szintű minták anélkül, hogy minden szinten az összes információt átmásolnánk a saját kéregállományunkba. Így a neokortexünkben minden egyes minta csak az alatta lévő szinteken hordozott összes információ fényében nyer értelmet. Ráadásul az azonos és magasabb szinteken lévő más minták is fontosak egy adott minta értelmezésében, mert ezek adják a kontextust. A valódi gondolatolvasáshoz tehát nem csupán a releváns axonok aktivációit kellene észlelni egy személy agyában, hanem lényegében az egész neokortexét kellene megvizsgálni, annak összes emlékével együtt, hogy megértsük ezeket az aktivációkat.

Ahogy megtapasztaljuk saját gondolatainkat és emlékeinket, "tudjuk", hogy mit jelentenek, de nem léteznek könnyen megmagyarázható gondolatok és emlékek formájában. Ha meg akarjuk osztani őket másokkal, le kell fordítanunk őket nyelvre. Ezt a feladatot is a neokortex végzi, olyan mintafelismerők segítségével, amelyeket a nyelvhasználat céljából megtanult mintákon edzettünk. A nyelv maga is erősen hierarchikus, és úgy fejlődött ki, hogy kihasználja a neokortex hierarchikus természetét, ami viszont a valóság hierarchikus természetét tükrözi. Az embereknek az a veleszületett képessége, hogy megtanulják a nyelv hierarchikus struktúráit, amelyekről Noam Chomsky írt, a neokortex szerkezetét tükrözi. Egy 2002-es tanulmányában, amelynek társszerzője volt, Chomsky a "rekurzió" attribútumát említi, mint ami az egyedi

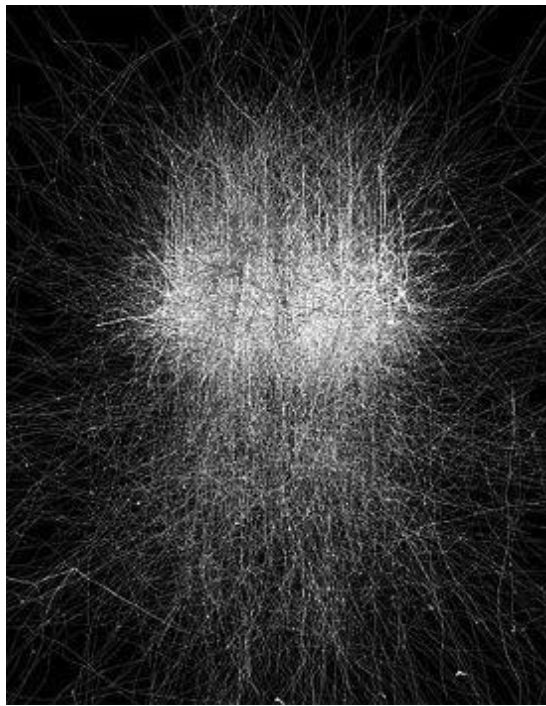
az emberi faj nyelvi képessége.⁴ Chomsky szerint a rekurzió az a képesség, hogy a kis részeket egy nagyobb darabkává rakjuk össze, majd ezt a darabkát egy másik szerkezet részeként használjuk fel, és ezt a folyamatot iteratív módon folytatjuk. Ily módon képesek vagyunk korlátozott számú szóból a mondatok és bekezdések bonyolult szerkezeteit felépíteni. Bár Chomsky itt nem kifejezetten az agy szerkezetére utalt, az általa leírt képesség pontosan az, amit a neokortex végez.

Az alacsonyabb rendű emlősfajok nagyrészt a neokortexüket használják fel a sajátos életmódjukból adódó kihívásokkal. Az emberi faj további képességekre tett szert azért, hogy lényegesen több agykéreg nőtt a beszélt és írott nyelv kezelésére. Egyes emberek jobban elsajátították ezeket a képességeket, mint mások. Ha sokszor elmeséltünk egy adott történetet, akkor ténylegesen elkezdjük megtanulni azt a nyelvi szekvenciát, amely a történetet különálló szekvenciák sorozataként írja le. A memóriánk még ebben az esetben

sem a szavak szigorúan vett sorrendje, hanem inkább olyan nyelvi struktúrák, amelyeket minden alkalommal, amikor a történetet előadjuk, konkrét szószorozatokká kell alakítanunk. Ezért van az, hogy egy történetet minden alkalommal, amikor megosztjuk, egy kicsit másképp mondunk el (hacsak nem tanuljuk meg a pontos szószorozatot mintaként).

A konkrét gondolkodási folyamatok minden egyes leírása esetében figyelembe kell vennünk a redundancia kérdését is. Mint említettem, nincs egyetlen olyan mintánk, amely az életünkben fontos entitásokat reprezentálná, függetlenül attól, hogy ezek az entitások érzékszervi kategóriákat, nyelvi fogalmakat vagy eseményekre vonatkozó emlékeket jelentenek. Minden fontos minta - minden szinten - sokszor ismétlődik. Ezen ismétlődések némelyike egyszerű ismétlést jelent, sok viszont különböző perspektívákat és nézőpontokat képvisel. Ez az egyik fő oka annak, hogy egy ismerős arcot különböző tájolásokból és különböző fényviszonyok mellett is fel tudunk ismerni. A hierarchia minden egyes szintje jelentős redundanciával rendelkezik, ami elegendő variabilitást tesz lehetővé, amely összhangban van az adott koncepcióval.

Ha tehát elképzeljük, hogy megvizsgáljuk az ön neokortexét, amikor egy bizonyos szeretett személyt néz, akkor a mintafelismerő axonok rengeteg tüzelését látnánk minden szinten, a primitív érzékszervi minták alapszintjétől kezdve egészen a szeretett személy képét reprezentáló sok különböző mintáig. A helyzet egyéb aspektusait, például az illető mozgását, beszédét és így tovább, reprezentáló tüzelések tömegét is látnánk. Ha tehát az élmény sokkal gazdagabbnak tűnik, mint egy egyszerű, a funkciók hierarchiáján felfelé vezető rendezett utazás, akkor az is.



Számítógépes szimuláció a neokortex sok egyidejű mintafelismerőjének működéséről.

A mintafelismerők hierarchiáján felfelé haladó alapvető mechanizmus azonban, amelyben minden magasabb fogalmi szint egy-egy absztraktabb és integráltabb fogalmat képvisel, továbbra is érvényes marad. Az információáramlás lefelé még nagyobb, mivel a felismert minta minden egyes aktivált szintje előrejelzéseket küld a következő alacsonyabb szintű mintafelismerőknek arról, hogy mivel fog valószínűleg legközelebb találkozni. Az emberi tapasztalat látszólagos bujasága annak a ténynek az eredménye, hogy a neokortexünkben lévő több százmillió mintafelismerő mindegyike egyidejűleg veszi figyelembe a bemeneti adatokat.

Az [5. fejezetben](#) az érintésből, látásból, hallásból és más érzékszervekből a neokortexbe jutó információáramlást fogom tárgyalni. Ezeket a korai bemeneteket olyan agykérgi régiók dolgozzák fel, amelyek a megfelelő típusú érzékszervi bemeneteknek vannak szentelve (bár e régiók hozzárendelésében óriási plaszticitás van, ami a neokortex alapvető egységes működését tükrözi). A fogalmi hierarchia a neokortex minden egyes érzékszervi régiójában a legmagasabb fogalmak felett folytatódik. Az agykérgi asszociációs területek integrálják a különböző szenzoros bemenetektől származó bemeneteket. Amikor hallunk valamit, ami talán úgy hangzik, mint a házastársunk hangja, majd látunk valamit, ami talán a jelenlétére utal, nem bonyolult logikai következtetési folyamatba kezdünk, hanem ezen érzékszervi felismerések kombinációjából azonnal érzékeljük, hogy a házastársunk jelen van. Az összes lényeges érzékszervi és észlelési jelet - talán még a

parfümjé vagy a kölnije illatát is - egyetlen többszintű érzékelésként integráljuk.

Az agykérgi érzékszervi asszociációs területek feletti fogalmi szinten még absztraktabb fogalmakkal is képesek vagyunk foglalkozni - érzékelni, emlékezni és gondolkodni róluk. A legmagasabb szinten olyan mintákat ismerünk fel, mint például *ez vicces*, vagy *ő csinos*, vagy *ez ironikus*, és így tovább. Emlékeink ezeket az absztrakt felismerési mintákat is tartalmazzák. Emlékezhetünk például arra, hogy sétáltunk valakivel, aki valami vicceset mondott, és mi neveltünk, bár magára a viccre nem biztos, hogy emlékszünk. Ennek az emlékezésnek az emléksorozata egyszerűen a humor érzékelését rögzítette, de nem a pontos tartalmát annak, ami vicces volt.

Az [előző fejezetben](#) megjegyeztem, hogy gyakran felismerünk egy mintát, még ha nem is ismerjük fel elég jól ahhoz, hogy le tudjuk írni. Például azt hiszem, hogy ki tudnám választani a babakocsis nő képét, akit ma korábban láttam a többi nő képeinek csoportjából, annak ellenére, hogy valójában nem tudom őt vizualizálni, és nem tudok sok konkrétumot leírni róla. Ebben az esetben az emlékezetem róla bizonyos magas szintű jellemzők listája. Ezekhez a vonásokhoz nem kapcsolódnak nyelvi vagy képi címkék, és nem pixelképek, így bár képes vagyok rá gondolni, nem tudom leírni. Ha azonban egy képet mutatnak róla, akkor fel tudom dolgozni a képet, ami ugyanazon magas szintű jellemzők felismerését eredményezi, amelyeket akkor ismertem fel, amikor először láttam őt. Ezáltal megállapíthatom, hogy a vonások egyeznek, és így magabiztosan kiválaszthatom a képét.

Bár csak egyszer láttam ezt a nőt a sétám során, valószínűleg már több másolata van a neokortexemben a mintájának. Ha azonban egy bizonyos ideig nem gondolok rá, akkor ezek a mintafelismerők átrendeződnek más mintákhoz. Ezért halványulnak el az emlékek az idő múlásával: A redundancia mennyisége csökken, amíg bizonyos emlékek ki nem hálnak. Most azonban, hogy emléket állítottam ennek a bizonyos nőnek azzal, hogy itt írtam róla, valószínűleg nem fogom olyan könnyen elfelejteni.

Autoasszociáció és invariancia

Az [előző fejezetben](#) tárgyaltam arról, hogyan ismerhetünk fel egy mintát akkor is, ha a teljes minta nincs jelen, és akkor is, ha az torz. Az első képességet autoasszociációnak nevezzük: az a képesség, hogy egy mintát önmagának egy részéhez társítunk. Minden mintafelismerő felépítése eredendően támogatja ezt a képességet.

Ahogy egy alacsonyabb szintű mintafelismerőből minden egyes bemenet felfelé áramlik egy magasabb szintű felismerőbe, a kapcsolatnak lehet egy "súlya", ami jelzi, hogy mennyire fontos az adott elem a mintában. Így a minta jelentősebb elemei nagyobb súlyt kapnak annak mérlegelése során, hogy az adott minta "felismerhető" legyen-e. Lincoln szakállának, Elvis pajeszának és Einstein híres nyelvmozdulatának valószínűleg nagy súlya van azokban a mintákban, amelyeket ezen ikonikus alakok megjelenéséről tanultunk. A mintafelismerő olyan valószínűséget számol, amely figyelembe veszi a fontossági paramétereket. Így az összvalószínűség alacsonyabb, ha egy vagy több elem hiányzik, bár a felismerési küszöbérték ennek ellenére teljesülhet. Amint rámutattam, az általános valószínűség (hogy a minta jelen van) kiszámítása bonyolultabb, mint egy egyszerű

súlyozott összeg, mivel a méretparamétereket is figyelembe kell venni.

Ha a mintafelismerő egy magasabb szintű felismerőtől azt a jelzést kapta, hogy a mintája "várható", akkor a küszöbértéket hatékonyan csökkentik (azaz könnyebben elérhetővé teszik). Alternatív megoldásként egy ilyen jel egyszerűen hozzáadódhat a súlyozott bemenetek összegéhez, így kompenzálva a hiányzó elemet. Ez minden szinten megtörténik, így egy olyan minta, mint például egy arc, amely több szinttel feljebb van, több hiányzó jellemzővel is felismerhető.

Azt a képességet, hogy a mintákat akkor is felismerjük, ha azok aspektusai átalakulnak, jellemzőinvariánsnak nevezzük, és négyféleképpen kezeljük. Először is, vannak olyan globális átalakítások, amelyek még azelőtt végbemennek, hogy a neokortex érzékszervi adatokat kapna. A szemtől, fültől és bőrtől érkező érzékszervi adatok útját a [94. oldal](#)on "[Az érzékszervi útvonal](#)" című részben tárgyaljuk.

A második módszer kihasználja az agykérgi mintamemóriánk redundanciáját. Különösen a fontos tárgyak esetében sok különböző perspektívát és nézőpontot tanultunk meg az egyes mintákhoz. Így sok variációt külön tárolunk és dolgozunk fel.

A harmadik és legerősebb módszer két lista kombinálásának lehetősége. Az egyik lista tartalmazhat egy sor olyan transzformációt, amelyekről megtanultuk, hogy alkalmazhatók egy bizonyos kategóriájú mintára; a kéreg ugyanezt a lehetséges változások listáját alkalmazza egy másik mintára. Így értjük meg az olyan nyelvi jelenségeket, mint a metaforák és hasonlatok.

Megtanultuk például, hogy bizonyos fonémák (a nyelv alaphangjai) hiányozhatnak a beszélt beszédből (például a "goi(n)"). Ha ezután megtanulunk egy új beszélt szót (például "vezetés"), akkor képesek leszünk felismerni azt a szót, ha az egyik fonéma hiányzik belőle, még akkor is, ha korábban soha nem tapasztaltuk ezt a szót ebben az alakban, mert megismerkedtünk azzal az általános jelenséggel, hogy bizonyos fonémák kimaradnak. Egy másik példaként megtanulhatjuk, hogy egy adott művész szereti kiemelni (nagyobbá tenni) egy arc bizonyos elemeit, például az orrát. Ilyenkor azonosítani tudunk egy általunk ismert arcot, amelyen ezt a módosítást alkalmazták, még akkor is, ha soha nem láttuk ezt a módosítást az adott arcon. Bizonyos művészi módosítások éppen azokat a vonásokat emelik ki, amelyeket a mintafelismerésen alapuló neokortexünk felismer. Mint említettük, pontosan ez a karikatúra alapja.

A negyedik módszer a méretparaméterekből ered, amelyek lehetővé teszik, hogy egyetlen modul egy minta több példányát kódolja. Például sokszor hallottuk már a "meredek" szót. Egy adott mintafelismerő modul, amely felismeri ezt a kimondott szót, kódolhatja ezt a többszörös példát azzal, hogy jelzi, hogy az [E] időtartama nagy várható változékonysággal rendelkezik. Ha az [E]-t tartalmazó szavak összes modulja hasonló jelenséget mutat, akkor ezt a változékonyságot magának az [E]-nek a modelljeiben is kódolhatjuk. Az [E]-t (vagy sok más fonémát) tartalmazó különböző szavak azonban eltérő mértékű várható változékonysággal rendelkezhetnek. Például a "csúcs" szóban valószínűleg nem lesz annyira kihúzva az [E] fonéma, mint a "meredek" szóban.

Tanulás

Nem mi magunk teremtjük-e utódainkat a föld feletti uralomban? Napról napra növeljük szervezetük szépségét és finomságát, napról napra nagyobb képességekkel ruházzuk fel őket, és egyre többet és többet adunk abból az önszabályozó, önműködő erőből, amely jobb lesz minden értelemnél?

-Samuel Butler, 1871

Az agyak fő tevékenységei a változások elvégzése önmagukban.

-Marvin Minsky, *Az elme társadalma*

Eddig azt vizsgáltuk, hogyan ismerünk fel (érzékszervi és észlelési) mintákat, és hogyan idézünk fel mintasorozatokat (a dolgokról, emberekről és eseményekről való emlékezetünk). Azonban nem úgy születünk, hogy a neokortexünk tele van ezekkel a mintákkal. A neokortexünk szűz terület, amikor agyunk létrejön. Képes a tanulásra, és ezért képes kapcsolatokat létrehozni a mintázatfelismerői között, de ezeket a kapcsolatokat a tapasztalatokból nyeri.

Ez a tanulási folyamat már születésünk előtt elkezdődik, és egyidejűleg zajlik az agy tényleges növekedésének biológiai folyamatával. A magzatnak már egy hónapos korában van agya, bár ez lényegében egy hüllőagy, mivel a magzat valójában a biológiai evolúció nagy sebességű újratereztésén megy keresztül az anyaméhben. A születési agy a terhesség harmadik trimeszterére már egyértelműen emberi agy, emberi neokortexszel. Ekkor a magzatnak már tapasztalatai vannak, és a neokortex tanul. Hallja a hangokat, különösen az anyja szívverését, ami az egyik valószínű oka annak, hogy a zene ritmikus tulajdonságai univerzálisak az emberi kultúrában. Minden valaha felfedezett emberi civilizáció kultúrájának része volt a zene, ami más művészeti ágakra, például a képzőművészetre nem jellemző. Az is igaz, hogy a zene ritmusa a szívritmusunkhoz hasonlítható. A zene üteme bizonyára változik - máskülönben a zene nem tartaná fenn az érdeklődésünket -, de a szívverés is változik. A túlságosan szabályos szívverés valójában a beteg szív tünete. A magzat szeme a fogantatás után huszonhat héttel már részben nyitva van, és a fogantatás után huszonnyolc héttel már többnyire teljesen nyitva van. Lehet, hogy az anyaméhben nem sok mindent lehet látni, de vannak a fény és a sötétség mintázatai, amelyeket a neokortex elkezd feldolgozni.

Tehát bár az újszülött csecsemőnek volt némi tapasztalata az anyaméhben, ez nyilvánvalóan korlátozott. A neokortex is tanulhat az öreg agytól (ezt a témát az [5. fejezet](#)ben tárgyalom), de általában véve a születéskor a gyermeknek sok mindent meg kell tanulnia - az alapvető primitív hangoktól és formáktól kezdve a metaforáig és a szarkazmusig mindent.

A tanulás kritikus fontosságú az emberi intelligencia szempontjából. Ha tökéletesen modelleznénk és szimulálnánk az emberi neokortexet (ahogyan azt a Blue Brain Project

próbálja megtegni) és az összes többi agyi régiót, amelyre a működéséhez szüksége van (például a hippokampuszt és a talamuszt), akkor nem sok mindenre lenne képes - ugyanúgy, ahogyan egy újszülött csecsemő sem tudna sokat tenni (azon kívül, hogy aranyos, ami határozottan kulcsfontosságú túlélési adaptáció).

A tanulás és a felismerés egyszerre történik. Azonnal elkezdünk tanulni, és amint megtanultunk egy mintát, azonnal elkezdjük felismerni azt. A neokortex folyamatosan próbálja értelmezni a neki bemutatott inputot. Ha egy adott szint nem képes teljes mértékben feldolgozni és felismerni egy mintát, akkor azt a következő, magasabb szintre továbbítja. Ha egyik szintnek sem sikerül felismerni egy mintát, akkor azt új mintának tekintjük. Egy minta újnak minősítése nem feltétlenül jelenti azt, hogy annak minden aspektusa új. Ha egy adott művész festményeit nézegetjük, és egy macska arcát látjuk egy elefánt orrával, akkor képesek leszünk azonosítani az egyes megkülönböztető jegyeket, de észrevesszük, hogy ez a kombinált minta valami újszerű, és valószínűleg emlékezni fogunk rá. A neokortex magasabb fogalmi szintjei, amelyek megértik a kontextust - például azt a körülményt, hogy ez a kép egy adott művész munkásságának példája, és hogy éppen az adott művész új festményeinek kiállításának megnyitóján vagyunk jelen -, észreveszik a macska-elefánt arc szokatlan mintakombinációját, de ezeket a kontextuális részleteket is további emlékezeti mintaként fogják felvenni.

Az új emlékek, mint például a macskaelefánt-arc, egy rendelkezésre álló mintafelismerőben tárolódnak. A hippokampusz szerepet játszik ebben a folyamatban, és a következő fejezetben tárgyaljuk, hogy mit tudunk a tényleges biológiai mechanizmusokról. Neokortex-modellünk szempontjából elegendő azt mondani, hogy az egyébként fel nem ismert mintákat új mintaként tároljuk, és megfelelően összekapcsoljuk az őket alkotó alacsonyabb szintű mintákkal. A macska-elefánt arcot például többféleképpen tároljuk el: Az arcrészek újszerű elrendezése ugyanúgy tárolásra kerül, mint a kontextuális emlékek, amelyek magukban foglalják a művészt, a helyzetet, és talán azt a tényt, hogy neveltünk, amikor először láttuk.

A sikeresen felismert emlékek egy új minta létrehozását is eredményezhetik a nagyobb redundancia elérése érdekében. Ha a mintákat nem ismerik fel tökéletesen, akkor valószínűleg úgy tárolják el őket, mint amelyek a felismert elem egy másik nézőpontját tükrözik.

Mi tehát az általános módszer annak meghatározására, hogy mely minták kerülnek tárolásra? Matematikailag a probléma a következőképpen fogalmazható meg: A rendelkezésre álló mintatárolási korlátokat kihasználva hogyan tudjuk optimálisan reprezentálni az eddig bemutatott bemeneti mintákat? Bár van értelme bizonyos mértékű redundanciát megengedni, nem lenne célszerű a teljes rendelkezésre álló tárolási területet (azaz az egész neokortexet) ismétlődő mintákkal megtölteni, mivel ez nem tenné lehetővé a minták megfelelő változatosságát. Egy olyan mintát, mint például az [E] fonéma a beszélt szavakban, számtalanszor tapasztaltunk már. Ez a hangfrekvenciák egyszerű mintázata, és kétségtelenül jelentős redundanciát élvez a neokortexünkben. Az egész neokortexünket meg tudnánk tölteni az [E] fonéma ismétlődő mintáival. A hasznos redundanciának azonban van egy határa, és egy ilyen gyakori minta, mint ez, egyértelműen elérte ezt a határt.

Erre az optimalizálási problémára létezik egy matematikai megoldás, az úgynevezett

lineáris programozás, amely a korlátozott erőforrások (ebben az esetben a mintafelismerők korlátozott száma) lehető legjobb elosztását oldja meg, amely az összes olyan esetet reprezentálja, amelyre a rendszer betanult. A lineáris programozást egydimenziós bemenetekkel rendelkező rendszerekhez tervezték, ami egy másik ok, amiért optimális az egyes mintafelismerő modulok bemenetét bemenetek lineáris sorozataként ábrázolni. Ezt a matematikai megközelítést használhatjuk egy szoftverrendszerben, és bár egy tényleges agyat tovább korlátoznak a rendelkezésére álló fizikai kapcsolatok, amelyekkel a mintafelismerők között alkalmazkodni tud, a módszer mégis hasonló.

Ennek az optimális megoldásnak fontos következménye, hogy a rutinszerű tapasztalatok felismerésre kerülnek, de nem eredményeznek maradandó emléket. A sétámat illetően milliónyi mintát tapasztaltam minden szinten, az alapvető vizuális élek és árnyalatoktól kezdve az olyan tárgyakig, mint a lámpaoszlopok és postaládák, valamint az emberek, állatok és növények, amelyek mellett elhaladtam. Szinte semmi sem volt egyedi abból, amit tapasztaltam, és az általam felismert minták már régen elérték a redundancia optimális szintjét. Az eredmény az, hogy szinte semmire sem emlékszem erről a sétáról. Azt a néhány részletet, amire emlékszem, valószínűleg felülírják az új minták, mire újabb néhány tucat sétát teszek - kivéve azt a tényt, hogy most emléket állítottam ennek a bizonyos sétának azzal, hogy írtam róla.

Az egyik fontos pont, amely mind a biológiai neokortexünkre, mind pedig az azt utánzó kísérletekre vonatkozik, hogy nehéz egyszerre túl sok fogalmi szintet megtanulni. Egyszerre alapvetően egy vagy legfeljebb két fogalmi szintet tudunk megtanulni. Amint ez a tanulás viszonylag stabil, továbbléphetünk a következő szint megtanulásához. Folytathatjuk a tanulás finomhangolását az alsóbb szinteken, de a tanulás fókusza a következő absztrakciós szintre irányul. Ez igaz mind az élet kezdetén, amikor az újszülöttek az alapvető formákkal küzdenek, mind pedig életünk későbbi szakaszában, amikor új tárgyak elsajátításával küzdünk, egyszerre csak egy-egy komplexitási szintet. Ugyanezt a jelenséget találjuk a neokortex gépi emulációiban is. Ha azonban egyre absztraktabb anyagot mutatnak be nekik szintről szintre, a gépek ugyanúgy képesek tanulni, mint az emberek (bár még nem olyan sok fogalmi szinten).

Egy minta kimenete visszahat egy alacsonyabb szintű mintára vagy akár magára a mintára is, ami az emberi agy erőteljes rekurzív képességét adja. Egy minta eleme lehet egy másik mintán alapuló döntési pont. Ez különösen hasznos olyan listák esetében, amelyek cselekvéseket állítanak össze - például egy másik tubus fogkrém beszerzése, ha az aktuális kiürült. Ezek a feltételesek minden szinten léteznek. Mint azt bárki tudja, aki próbált már eljárást programozni a számítógépen, a feltételes függvények létfontosságúak egy műveletmenet leírásához.

A gondolkodás nyelve

Az álom biztonsági szelepként működik a túlterhelt agy számára.

-Sigmund Freud,

Az álmok értelmezése, 1911

Agy: az a készülék, amellyel gondolkodunk, gondolkodunk.

-Ambrose Bierce, *Az ördög szótára*

A neokortex működéséről eddig tanultakat összefoglalva lásd a neokortex mintafelismerő moduljának ábráját a [42.](#) oldalon.

a) A dendritek belépnek a mintát reprezentáló modulba. Bár a minták két- vagy háromdimenziósak tűnhetnek, a jelek egydimenziós sorozatával reprezentálódnak. A mintának ebben a (szekvenciális) sorrendben kell jelen lennie ahhoz, hogy a mintafelismerő képes legyen felismerni azt. A dendritek mindegyike végső soron egy vagy több, alacsonyabb fogalmi szinten lévő mintafelismerő axonjához kapcsolódik, amelyek egy alacsonyabb szintű mintát ismertek fel, amely ennek a mintának a részét képezi. Minden egyes ilyen bemeneti minta esetében számos alacsonyabb szintű mintafelismerő lehet, amelyek képesek létrehozni azt a jelet, hogy az alacsonyabb szintű mintát felismerték. A minta felismeréséhez szükséges küszöbértéket akkor is el lehet érni, ha nem minden bemenet jelzett. A modul kiszámítja annak a valószínűségét, hogy az általa felismert minta jelen van. Ez a számítás figyelembe veszi a "fontosság" és a "méret" paramétereit (lásd alább [f]).

Vegyük észre, hogy a dendritek egy része jeleket továbbít a modulba, más része pedig a modulból kifelé. Ha a mintafelismerőhöz bejövő összes dendrit jelzi, hogy egy vagy két kivétellel az alacsonyabb szintű mintákat már felismerték, akkor ez a mintafelismerő jelet küld lefelé a még fel nem ismert alacsonyabb szintű mintákat felismerő mintafelismerő(k)nek, jelezve, hogy nagy a valószínűsége, hogy az adott mintát hamarosan felismerik, és az alacsonyabb szintű mintafelismerő(k)nek résen kell lenniük.

b) Amikor ez a mintafelismerő felismeri a mintáját (a bemeneti dendritek összes vagy legtöbb jelének aktiválódása alapján), a mintafelismerő axonja (kimenete) aktiválódik. Ez az axon viszont egy egész dendrithálózathoz kapcsolódhat, amely számos magasabb szintű mintafelismerőhöz kapcsolódik, amelyeknek ez a minta a bemenete. Ez a jel nagyságrendi információt továbbít, hogy a következő, magasabb fogalmi szinten lévő mintafelismerők figyelembe vehessék azt.

- c) Ha egy magasabb szintű mintafelismerő pozitív jelet kap az összes vagy a legtöbb alkotó mintától, kivéve az adott mintafelismerő által képviselt mintát, akkor a magasabb szintű felismerő leküldhet egy jelet ennek a felismerőnek, jelezve, hogy az ő mintája várható. Egy ilyen jel hatására ez a mintafelismerő csökkentené a küszöbértékét, ami azt jelenti, hogy nagyobb valószínűséggel küldene jelet az axonján (jelezve, hogy a mintáját felismertnek tekintik) még akkor is, ha néhány bemenete hiányzik vagy nem egyértelmű.
- d) Az alulról érkező gátló jelek kevésbé valószínűvé tennék, hogy ez a mintafelismerő felismeri a mintáját. Ez következhet olyan alacsonyabb szintű minták felismeréséből, amelyek nem állnak összhangban az adott mintafelismerőhöz társított mintával (például egy bajusz felismerése egy alacsonyabb szintű felismerő által kevésbé valószínűvé tenné, hogy ez a kép "a feleségem").
- e) A felülről érkező gátló jelek szintén csökkentik annak valószínűségét, hogy ez a mintafelismerő felismeri a mintáját. Ez egy olyan magasabb szintű kontextusból adódhat, amely nincs összhangban az adott felismerőhöz tartozó mintával.
- f) Minden egyes bemenethez tárolt paraméterek vannak a fontosságra, a várható méretre és a méret várható változékonyságára vonatkozóan. A modul kiszámítja annak általános valószínűségét, hogy a minta jelen van, mindezen paraméterek és az aktuális jelek alapján, amelyek jelzik, hogy a bemenetek közül melyek vannak jelen, és azok nagysága. Ennek matematikailag optimális módja a rejtett Markov-modelleknek nevezett technika. Ha az ilyen modellek hierarchiába szerveződnek (mint a neokortexben vagy a neokortex szimulálására tett kísérletekben), hierarchikus rejtett Markov-modelleknek nevezzük őket.

A neokortexben kiváltott minták más mintákat váltanak ki. A részben befejezett minták jeleket küldenek lefelé a fogalmi hierarchiában; a befejezett minták jeleket küldenek felfelé a fogalmi hierarchiában. Ezek a neokortikális minták a gondolkodás nyelve. A nyelvhez hasonlóan hierarchikusak, de önmagukban nem nyelv. Gondolataink nem elsősorban a nyelv elemeiben fogalmazódnak meg, bár mivel a nyelv is mint minták hierarchiájaként létezik a neokortexünkben, lehetnek nyelvi alapú gondolataink. De a gondolatok nagyrészt ezekben a neokortikális mintákban reprezentálódnak.

Ahogy fentebb már említettem, ha képesek lennénk észlelni a mintaaktivációkat valakinek a neokortexében, akkor is kevés fogalmunk lenne arról, hogy ezek a mintaaktivációk mit jelentenek anélkül, hogy hozzáférnénk az egyes aktivált minták feletti és alatti minták teljes hierarchiájához. Ehhez nagyjából az illető teljes neokortexéhez kellene hozzáférnünk. Elég nehéz számunkra megérteni a saját gondolataink tartalmát, de egy másik

ember gondolatainak megértéséhez egy, a miénktől eltérő neokortex elsajátítására van szükség. Természetesen még nincs hozzáférésünk valaki más neokortexéhez; ehelyett arra kell hagyatkoznunk, hogy megpróbálja a gondolatait nyelvben kifejezni (valamint más eszközökkel, például gesztusokkal). Az emberek hiányos képessége ezeknek a kommunikációs feladatoknak a teljesítésére egy újabb komplexitási réteget ad hozzá - nem csoda, hogy ennyire félreértjük egymást.

Kétféle gondolkodásmódunk van. Az egyik a nem irányított gondolkodás, amelyben a gondolatok nem logikus módon váltják ki egymást. Amikor valami más tevékenység közben, például a levelek gereblyézése vagy az utcán való sétálás közben hirtelen eszünkbe jut egy évvel vagy évtizedekkel ezelőtti emlék, az élményt - mint minden emléket - minták sorozataként idézzük fel. Nem képzeljük el azonnal a jelenetet, hacsak nem tudunk sok más emléket előhívni, amelyek lehetővé teszik számunkra, hogy egy erőteljesebb visszaemlékezést szintetizáljunk. Ha ily módon mégis vizualizáljuk a jelenetet, akkor lényegében az emlékezés pillanatában lévő utalásokból hozzuk létre az elménkben; maga az emlék nem képek vagy vizualizációk formájában tárolódik. Ahogy korábban említettem, a kiváltó okok, amelyek miatt ez a gondolat felbukkant az elménkben, vagy nyilvánvalóak, vagy nem. A vonatkozó gondolatok sorozata lehet, hogy azonnal elfelejtődött. Még ha emlékszünk is rá, az asszociációk nem lineáris és körkörös sorozata lesz.

A második gondolkodási mód az irányított gondolkodás, amelyet akkor használunk, amikor megpróbálunk megoldani egy problémát vagy megfogalmazni egy szervezett választ. Például elpróbálhatunk fejben valamit, amit mondani akarunk valakinek, vagy megfogalmazhatunk egy szöveget, amit meg akarunk írni (esetleg egy elmés könyvben). Ahogy az ilyen és ehhez hasonló feladatokra gondolunk, már mindegyik feladatot részfeladatok hierarchiájára bontottuk. Egy könyv megírása például fejezetek írását jelenti; minden fejezetnek vannak szakaszai; minden szakasznak vannak bekezdései; minden bekezdés tartalmaz gondolatokat kifejező mondatokat; minden gondolatnak megvan az elemekből álló konfigurációja; minden elem és az elemek közötti minden kapcsolat egy gondolat, amelyet meg kell fogalmazni; és így tovább. Ugyanakkor neokortikális struktúráink megtanultak bizonyos szabályokat, amelyeket követni kell. Ha a feladat az írás, akkor meg kell próbálnunk elkerülni a felesleges ismétléseket; meg kell próbálnunk biztosítani, hogy az olvasó követni tudja, amit írunk; meg kell próbálnunk betartani a nyelvtani és stílusbeli szabályokat; és így tovább. Az írónak tehát fel kell építenie az olvasó modelljét a fejében, és ez a konstrukció is hierarchikus. Az irányított gondolkodás során neokortexünkben listákat lépegetünk végig, amelyek mindegyike allisták kiterjedt hierarchiájává bővül, mindegyiknek saját szempontjai vannak. Ne feledjük, hogy a neokortikális mintázatban lévő lista elemei tartalmazhatnak feltételeseket, így a későbbi gondolataink és cselekedeteink a folyamat során elvégzett értékeléseken fognak múlni.

Sőt, minden ilyen irányított gondolat irányítatlan gondolatok hierarchiáját váltja ki. A töprengések folyamatos vihara kíséri mind az érzékszervi tapasztalatainkat, mind az irányított gondolkodásra tett kísérleteinket. Tényleges mentális tapasztalatunk összetett és kusza, a kiváltott minták e villámviharaiból áll, amelyek másodpercenként körülbelül százszor változnak.

Az álmok nyelve

Az álmok a nem irányított gondolatok példái. Van némi értelmük, mert az a jelenség, hogy az egyik gondolat kiváltja a másikat, a neokortexünkben lévő minták tényleges kapcsolatain alapul. Amennyiben egy álomnak nincs értelme, azt a konfabulációs képességünkkel próbáljuk helyrehozni. Amint azt a [9. fejezet](#)ben leírom, a hasadt agyú betegek (akiknek az agy két féltekéjét összekötő corpus callosum átvágódott vagy megsérült) a bal agyféltekéjükkel - amely a beszédközpontot irányítja - konfabulálnak (magyarázatokot találnak ki), hogy megmagyarázzák, amit a jobb agyfélteke éppen csinált olyan inputokkal, amelyekhez a bal agyfélteke nem férhetett hozzá. Folyton konfabulálunk, amikor események kimenetelét magyarázzuk. Ha jó példát akarnak erre, csak hallgassák a pénzügyi piacok mozgásáról szóló napi kommentárokat. Nem számít, hogyan teljesítenek a piacok, mindig lehet egy jó magyarázatot találni arra, hogy miért történt, és ilyen utólagos kommentárok bőven akadnak. Természetesen, ha ezek a kommentátorok valóban értenék a piacokat, nem kellene az idejüket kommentálásra pazarolniuk.

A konfabulálás természetesen a neokortexben is zajlik, amely jól tud olyan történeteket és magyarázatokat kitalálni, amelyek megfelelnek bizonyos korlátoknak. Ezt tesszük, amikor újra elmesélünk egy történetet. Olyan részleteket töltünk ki, amelyek esetleg nem állnak rendelkezésünkre, vagy amelyeket elfelejtettünk, hogy a történetnek több értelme legyen. Ezért változnak a történetek az idő múlásával, amikor új és új mesélők mesélik el őket újra és újra, esetleg más-más céllal. Ahogy azonban a beszélt nyelv átváltott az írott nyelvre, olyan technológiával rendelkezünk, amely képes volt rögzíteni egy történet végeleges változatát, és megakadályozni az ilyenfajta sodródást.

Az álom tényleges tartalma, amennyiben emlékszünk rá, ismét minták sorozata. Ezek a minták egy történet kényszereit jelentik; mi aztán egy olyan történetet konfabulálunk, amely megfelel ezeknek a kényszereknek. Az álomnak az a változata, amelyet újra elmesélünk (még ha csak magunknak némán is), ez a konfabuláció. Ahogy elmeséljük az álmot, olyan minták kaszkádját indítjuk el, amelyek kitöltik a tényleges álmot, ahogyan eredetileg átéltük.

Van egy alapvető különbség az álomgondolatok és az ébrenléti gondolkodásunk között. Az egyik lecke, amit az életben megtanulunk, hogy bizonyos cselekedetek, sőt gondolatok sem megengedettek a való világban. Megtanuljuk például, hogy nem teljesíthetjük azonnal a vágyainkat. Vannak szabályok, amelyek tiltják, hogy a bolt pénztárgépében lévő pénzt megragadjuk, és korlátok vannak arra vonatkozóan, hogy kapcsolatba lépjünk egy olyan személlyel, akihez fizikailag vonzódunk. Azt is megtanuljuk, hogy bizonyos gondolatok nem megengedettek, mert kulturálisan tiltottak. A szakmai készségek elsajátításával megtanuljuk a szakmánkban elismert és jutalmazott gondolkodásmódokat, és ezáltal elkerüljük azokat a gondolkodási mintákat, amelyek elárulhatják az adott szakma módszereit és normáit. Sok ilyen tabu érdemes, mivel érvényre juttatja a társadalmi rendet és megszilárdítja a haladást. Ugyanakkor megakadályozhatják a fejlődést is, mivel egy terméketlen ortodoxiát kényszerítenek ki. Pontosan ilyen ortodoxiát hagyott maga után Einstein, amikor gondolatkísérleteivel megpróbálta meglovagolni a

fénysugarat.

A kulturális szabályokat a neokortexben a régi agy, különösen az amygdala segítségével érvényesítik. Minden gondolatunk más gondolatokat vált ki, és ezek közül néhány a kapcsolódó veszélyekkel kapcsolatos. Megtanuljuk például, hogy egy kulturális norma megszegése még a privát gondolatainkban is kiközösítéshez vezethet, amiről a neokortex felismeri, hogy veszélyezteti a jólétünket. Ha ilyen gondolatokkal foglalkozunk, az amygdala működésbe lép, és ez félelmet generál, ami általában az adott gondolat megszüntetéséhez vezet.

Az álmokban azonban ezek a tabuk lazulnak, és gyakran álmodunk olyan dolgokról, amelyek kulturálisan, szexuálisan vagy szakmailag tiltottak. Mintha az agyunk rájönne, hogy álmodás közben nem vagyunk tényleges szereplői a világnak. Freud írt erről a jelenségről, de azt is megjegyezte, hogy az ilyen veszélyes gondolatokat leplezzük, legalábbis amikor megpróbáljuk felidézni őket, hogy az éber agy továbbra is védve legyen tőlük.

A szakmai tabuk feloldása hasznosnak bizonyul a kreatív problémamegoldás szempontjából. Minden este alkalmazok egy mentális technikát, amikor lefekvés előtt egy adott problémán gondolkodom. Ez olyan gondolatsorokat indít el, amelyek az álmaimban is folytatódnak. Amint álmodom, a probléma megoldására a napközbeni szakmai korlátok terhe nélkül tudok gondolkodni-álmodni a megoldásokról. Ezekhez az álmogondolatokhoz aztán reggel, az álmodás és az ébrenlét köztes állapotában férhetek hozzá, néha az úgynevezett "világos álmodás".⁵

Freud arról is híres, hogy az álmok értelmezésével betekintést nyerhetünk egy személy pszichológiájába. Ennek az elméletnek természetesen hatalmas irodalma van minden aspektusáról, de az alapvető elképzelés, hogy az álmaink vizsgálatán keresztül betekintést nyerhetünk önmagunkba, értelmes. Álmainkat a neokortexünk hozza létre, és így azok tartalma árulkodó lehet az ott található tartalmakról és összefüggésekről. Az ébrenlétünkben fennálló gondolkodásunk korlátozásainak lazulása szintén hasznos a neokortikális tartalmak feltárásában, amelyekhez egyébként nem tudnánk közvetlenül hozzáférni. Az is ésszerű következtetés, hogy az álmainkban megjelenő minták számunkra fontos dolgokat képviselnek, és ezáltal nyomokat adnak megoldatlan vágyaink és félelmeink megértéséhez.

A modell gyökerei

Mint már említettem, az 1980-as és 1990-es években egy olyan csoportot vezettem, amely kifejlesztette a hierarchikus rejtett Markov-modellek technikáját az emberi beszéd felismerésére és a természetes nyelvű kijelentések megértésére. Ez a munka volt az elődje a ma széles körben elterjedt kereskedelmi rendszereknek, amelyek felismerik és megértik, amit mondani akarunk (autós navigációs rendszerek, amelyekhez beszélni lehet, Siri az iPhone-on, Google Voice Search és sok más). Az általunk kifejlesztett technika lényegében

az összes olyan tulajdonsággal rendelkezett, amelyet a PRTM-ben leírtam. Tartalmazta a minták hierarchiáját, ahol minden magasabb szint fogalmilag absztraktabb volt, mint az alatta lévő. Például a beszédfelismerésben a szintek a legalacsonyabb szinten a hangfrekvencia alapmintáit, majd a fonémákat, végül a szavakat és a kifejezéseket (amelyeket gyakran úgy ismertek fel, mintha szavak lennének). Néhány beszédfelismerő rendszerünk képes volt megérteni a természetes nyelvi parancsok jelentését, így még magasabb szinteken olyan szerkezetek szerepeltek, mint a főnév- és igekötők. Minden mintafelismerő modul egy alacsonyabb fogalmi szintről származó minták lineáris sorozatát tudta felismerni. Minden bemenetnek voltak paraméterei a fontosságra, a méretre és a méret változékonyságára vonatkozóan. Voltak "lefelé irányuló" jelzések, amelyek azt jelezték, hogy egy alacsonyabb szintű minta várható. Ezt a kutatást részletesebben a [7. fejezetben](#) tárgyalom.

Jeff Hawkins és Dileep George, a 2003PalmPilot2004, feltalálója és Dileep George kifejlesztett egy hierarchikus agykérgi modellt, amelyet hierarchikus időbeli memóriának neveznek. Sandra Blakeslee tudományos íróővel Hawkins ékesszólóan leírta ezt a modellt az *On Intelligence* című könyvében. Hawkins határozott érveket hoz fel a kérgi algoritmus egységessége és hierarchikus, lista alapú szerveződése mellett. Van néhány fontos különbség az *On Intelligence* című könyvben bemutatott modell és az általam ebben a könyvben bemutatott modell között. Ahogy a neve is jelzi, Hawkins az alkotólisták időbeli (időalapú) jellegét hangsúlyozza. Más szóval, a listák iránya mindig előrefelé mutat az időben. Magyarázata arra, hogy egy kétdimenziós mintázat, például a nyomtatott "A" betű jellemzőinek hogyan van időbeli iránya, a szemmozgásra épül. Megmagyarázza, hogy a képeket szakkádok segítségével vizualizáljuk, amelyek a szem nagyon gyors mozgásai, amelyeknek nem vagyunk tudatában. A neokortexbe érkező információ tehát nem kétdimenziós jellemzők halmaza, hanem inkább egy időben rendezett lista. Bár igaz, hogy a szemünk valóban nagyon gyors mozgásokat végez, a sorrend, amelyben egy minta, például az "A" betű jellemzőit látja, nem mindig következetes időbeli sorrendben történik. (Például a szem szakkádjai nem mindig regisztrálják az "A" betű felső csúcsát az alsó homorúsága előtt.) Ráadásul olyan vizuális mintát is felismerhetünk, amely csak néhány tíz milliszekundumig van jelen, ami túl rövid idő ahhoz, hogy a szemszakkádok beolvassák azt. Igaz, hogy a neokortexben lévő mintafelismerők egy mintát listaként tárolnak, és a lista valóban rendezett, de a sorrend nem feltétlenül az időt jelenti. Gyakran valóban ez a helyzet, de jelenthet térbeli vagy magasabb szintű fogalmi rendezést is, ahogyan azt fentebb tárgyaltam.

A legfontosabb különbség az a paraméterkészlet, amelyet a mintafelismerő modul minden egyes bemenetéhez felvettem, különösen a méret és a méretváltozékonyság paraméterei. Az 1980-as években tulajdonképpen ilyen típusú információk nélkül próbáltuk felismerni az emberi beszédet. Ezt az indokolta, hogy a nyelvészek azt mondták, hogy az időtartam-információ nem különösebben fontos. Ezt a szemléletet illusztrálják azok a szótárak, amelyek az egyes szavak kiejtését fonémák sorozataként írják ki, például az "meredek" szót [s]

[t] [E] [p], anélkül, hogy jeleznénk, hogy az egyes fonémák várhatóan meddig tartanak. Ebből az következik, hogy ha programokat hozunk létre a fonémák felismerésére, és aztán

találkozunk ezzel a négy fonéma sorozatával (egy beszélt mondatban), akkor képesnek kell lennünk felismerni azt a beszélt szót. Az a rendszer, amelyet ezzel a megközelítéssel építettünk, bizonyos mértékig működött, de nem elég jól ahhoz, hogy olyan jellemzőkkel is megbirkózzon, mint a nagy szókincs, a több beszélő és a szünet nélkül, folyamatosan beszélt szavak. Amikor a hierarchikus rejtett Markov-modellek technikáját használtuk az egyes bemenetek nagyságeloszlásának beépítésére, a teljesítmény az egekbe szökött.

FEJEZET 4 A BIOLÓGIAI NEOKORTEX

Mivel a fontos dolgok egy tokba kerülnek, van egy koponya az agyadnak, egy műanyag tok a fésűdnek és egy pénztárca a pénzednek.

-George Costanza, a *Seinfeld* "A fordított kukucskáló" epizódjában

Most először figyelhetjük meg az agy működését globálisan, olyan tisztán, hogy képesek leszünk felfedezni a csodálatos képességei mögött meghúzódó átfogó programokat.

-J. G. Taylor, B. Horwitz, és K. J. Friston

Röviden, az elme úgy dolgozik a kapott adatokon, mint a szobrász a kőtömbjén. Bizonyos értelemben a szobor az örökkévalóságtól fogva ott állt. De volt mellette ezer másféle, és egyedül a szobrásznak köszönhetjük, hogy ezt az egyet kiemelte a többi közül. Éppen így mindannyiunk világa, bármennyire is különbözik a róla alkotott többféle képünk, mindannyian az érzetek őskáoszába ágyazódva feküdt, amely közömbösen adta a puszta anyagot mindannyiunk gondolkodásának. Ha akarjuk, érvelésünkkel visszatekerhetjük a dolgokat a térnek és a nyüzsgő atomok mozgó felhőinek abba a fekete és tagolatlan folytonosságába, amelyet a tudomány az egyetlen valóságos világnak nevez. De mindvégig az a világ, *amelyet* érzünk és amelyben élünk, az lesz, amit őseink és mi, lassan halmozódó választási mozdulatokkal, mint a szobrászok, egyszerűen az adott anyag bizonyos részeinek elutasításával kihúztunk belőle. Más szobrászok, más szobrok ugyanabból a kőből! Más elmék, más világok ugyanabból az egyhangú és kifejezéstelen káoszából! Az én világom csak egy a millióból, egyformán beágyazott, egyformán valóságos azok számára, akik elvonatkoztatathatják őket. Mennyire más lehet a világ a hangya, a tintahal vagy a rák tudatában!

-William James

Az intelligencia a biológiai evolúció célja, vagy egyáltalán célja? Steven Pinker azt írja: "Az agyunkkal kapcsolatban soviniztikusak vagyunk, és azt gondoljuk, hogy az evolúció célja az agyunk."¹ majd azzal érvel, hogy "ennek semmi értelme..... A természetes szelekció közel sem törekszik az intelligenciára. A folyamatot a különbségek vezérlik a szaporodó szervezetek túlélési és szaporodási aránya egy adott környezetben. Idővel az organizmusok olyan terveket sajátítanak el, amelyek az adott környezetben való túléléshez és szaporodáshoz igazítják őket, pont; semmi sem húzza őket más irányba, mint az ott és akkor elért siker." Pinker arra a következtetésre jut, hogy "az élet egy sűrűn elágazó bokor,

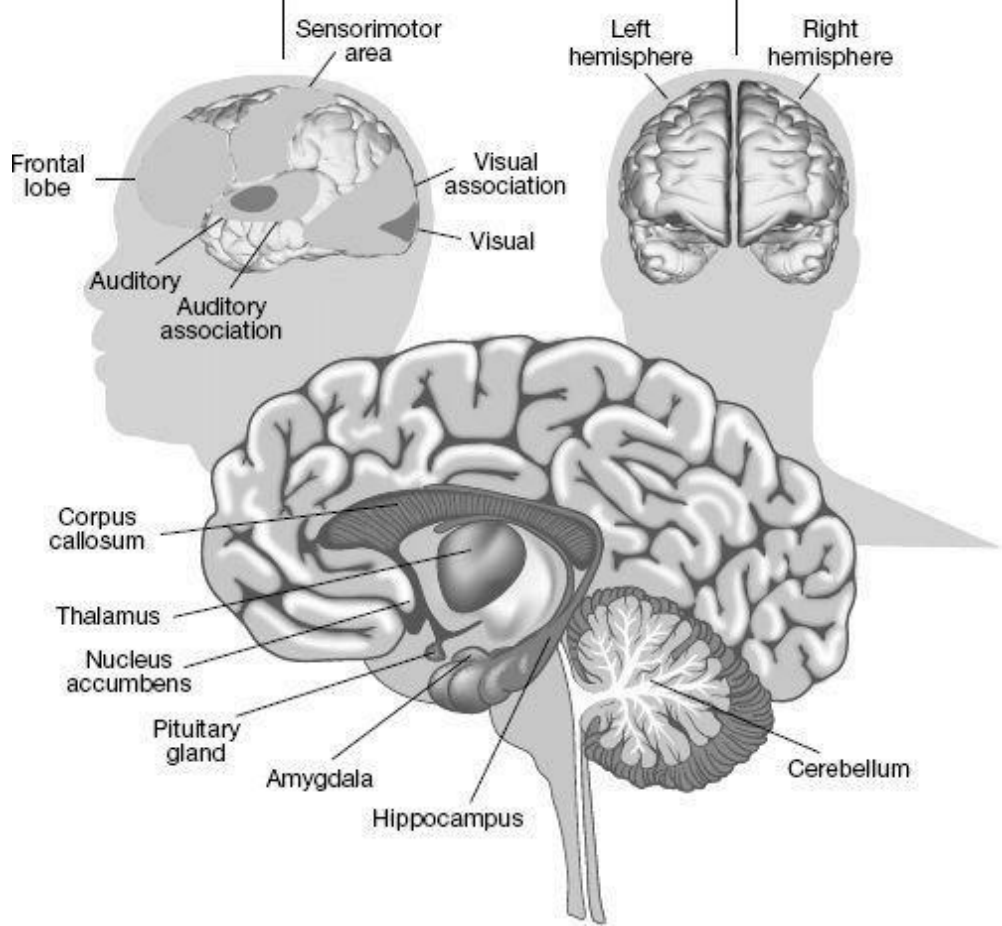
nem pedig egy skála vagy létra, és az élő szervezetek az ágak csúcán vannak, nem pedig az alsóbb fokokon".

Az emberi agyat illetően megkérdőjelezi, hogy "az előnyök meghaladják-e a költségeket". A költségek között említi, hogy "az agy terjedelmes. A női medence alig fér el egy csecsemő túlméretezett feje mellett. Ez a tervezési kompromisszum sok nőt megöl a szülés során, és olyan pendülő járást tesz szükségessé, amely a nőket biomechanikailag kevésbé hatékony járókelőkké teszi, mint a férfiakat. Emellett a nyakunkon billegő nehéz fej kiszolgáltatottabbá tesz minket a balesetek, például esések során bekövetkező halálos sérüléseknek." A továbbiakban további hiányosságokat sorol fel, köztük az agy energiafogyasztását, a lassú reakcióidőt és a tanulás hosszadalmas folyamatát.

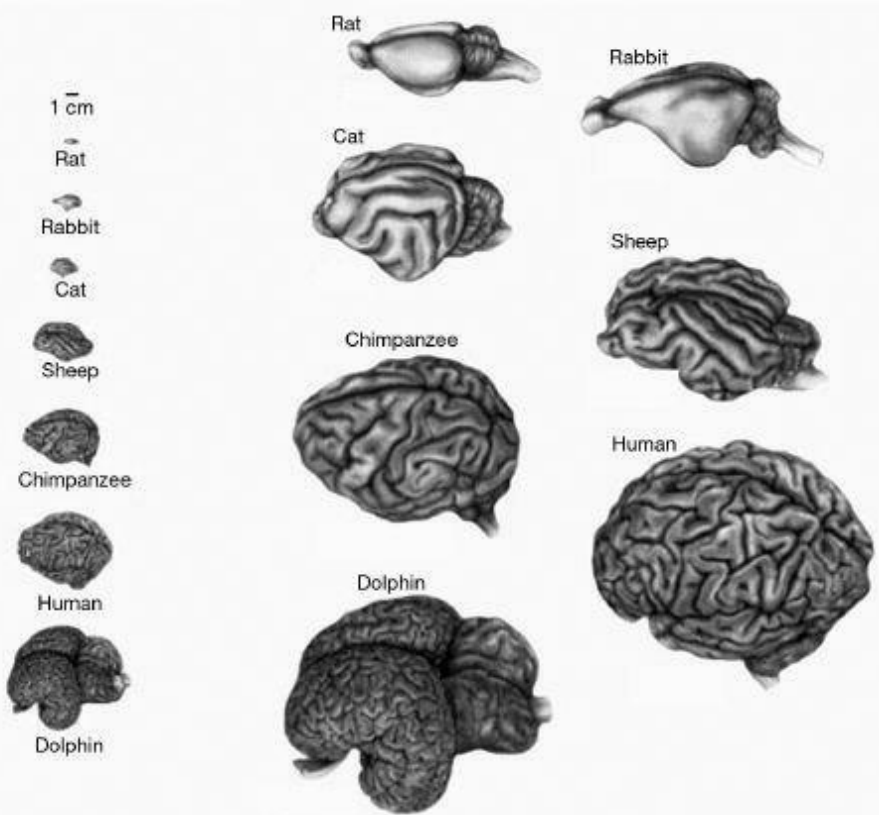
Bár ezek a kijelentések mindegyike helytálló (bár sok női barátom jobb gyalogló, mint én), Pinker nem érti a lényegét. Igaz, hogy biológiailag az evolúciónak nincs konkrét iránya. Ez egy keresési módszer, amely valóban alaposan kitölti a természet "sűrűn elágazó bozótját". Ugyanígy igaz az is, hogy az evolúciós változások nem *feltétlenül* a nagyobb intelligencia irányába haladnak - *minden* irányba haladnak. Számos példa van sikeres élőlényekre, amelyek évmilliókon át viszonylag változatlanok maradtak. (Az alligátorok például 200 millió évre nyúlnak vissza, és sok mikroorganizmus ennél jóval régebbre nyúlik vissza). De a számtalan evolúciós ág alapos kiterjedése során az egyik irány, amelyikben mozog, az a nagyobb intelligencia irányába mutat. E vita szempontjából ez a lényeges pont.

Neocortex

The neocortex covers the entire brain with its convoluted thin surface and is separated in two halves, connected by the corpus callosum.



Az agy kulcsfontosságú régióinak fizikai elrendezése.



A neokortex különböző emlősöknél.

Tegyük fel, hogy van egy kék gáz egy üvegben. Amikor levesszük a fedelet, nincs olyan üzenet, amely a gáz összes molekulájának azt üzeni: "Hé, srácok, a fedő lekerült az üvegről, menjünk fel a nyílás felé és ki a szabadságba!". A molekulák csak azt csinálják, amit mindig is csináltak, vagyis látszólag irány nélkül mozognak minden irányba. De közben néhányan a tetejéhez közeledve valóban kimozdulnak az üvegből, és idővel a legtöbbjük követni fogja őket. Amint a biológiai evolúció rábukkant egy hierarchikus tanulásra képes neurális mechanizmusra, rájött, hogy az rendkívül hasznos az evolúció egyetlen célja, a túlélés szempontjából. A neokortex előnye akkor vált nyilvánvalóvá, amikor a gyorsan változó körülmények a gyors tanulásnak kedveztek. Mindenféle faj - növények és állatok - képesek megtanulni, hogy idővel alkalmazkodjanak a változó körülményekhez, de neokortex nélkül a genetikai evolúció folyamatát kell alkalmazniuk. Nagyon sok generációra - akár több ezer évre - is szükség lehet ahhoz, hogy egy neokortex nélküli faj jelentős új viselkedésmódokat (vagy a növények esetében más alkalmazkodási stratégiákat) tanuljon meg. A neokortex kiemelkedő túlélési előnye az volt, hogy napok alatt képes volt tanulni. Ha egy faj drámaian megváltozott körülményekkel találkozik, és a faj egyik tagja kitalál, felfedez vagy egyszerűen csak rábukkan (ez a három módszer mind az innováció változata) egy módszert, amellyel alkalmazkodni tud a változáshoz, a többi egyed észreveszi, megtanulja és lemásolja ezt a módszert, és az gyorsan víruszerűen elterjed az

egész populációban. A körülbelül 65 millióéval előtti kataklizmus kréta-paleogén kihalási esemény sok olyan, nem neokortexet hordozó faj gyors pusztulásához vezetett, amelyek nem tudtak elég gyorsan alkalmazkodni a hirtelen megváltozott környezethez. Ez jelentette a fordulópontot a neokortexet hordozó emlősök számára, hogy átvegyék ökológiai résüket. Így a biológiai evolúció úgy találta, hogy a neokortex hierarchikus tanulása olyan értékes, hogy az agynak ez a régiója tovább növekedett, míg nem gyakorlatilag átvette a *Homo sapiens* agyát.

Az idegtudomány felfedezései meggyőzően bizonyították a neokortex hierarchikus képességeinek kulcsszerepét, valamint bizonyítékot szolgáltatottak az elme mintafelismerési elméletére (PRTM). Ezek a bizonyítékok számos megfigyelés és elemzés között oszlanak meg, amelyek egy részét itt áttekintem. Donald O. Hebb (1904-1985) kanadai pszichológus tett egy első kísérletet a tanulás neurológiai alapjainak magyarázatára. 1949-ben leírt egy olyan mechanizmust, amelyben a neuronok fiziológiailag megváltoznak a tapasztalataik alapján, és ezzel megteremtette a tanulás és az agyi plaszticitás alapját: "Tegyük fel, hogy egy visszhangzó tevékenység (vagy "nyom") fennmaradása vagy ismétlődése hajlamos tartós sejtes változásokat előidézni, amelyek hozzáadódnak a stabilitásához..... Amikor az *A* sejt axonja elég közel van ahhoz, hogy egy *B* sejtet gerjesszen, és ismételten vagy tartósan részt vesz annak tüzelésében, akkor az egyik vagy mindkét sejtben valamilyen növekedési folyamat vagy anyagcsere-változás megy végbe, oly módon, hogy *A* mint a *B*-t tüzelő sejtek egyike hatékonysága megnő."² Ezt az elméletet úgy fogalmazták meg, hogy "az együtt tüzelő sejtek

összekötve", és Hebbian tanulás néven vált ismertté. Hebb elméletének egyes aspektusai megerősítést nyertek, mivel egyértelmű, hogy az agyi egységek saját aktivitásuk alapján képesek új kapcsolatokat létrehozni és megerősíteni. Az agyi felvételeken valóban láthatjuk, hogy a neuronok ilyen kapcsolatokat alakítanak ki. A mesterséges "neurális hálók" a neuronális tanulás Hebb-féle modelljén alapulnak.

Hebb elméletének központi feltételezése, hogy a neokortexben a tanulás alapegysége a neuron. Az elme mintafelismerési elmélete, amelyet ebben a könyvben kifejték, egy másik alapegységen alapul: nem magán a neuronon, hanem inkább neuronok összességén, amelyek száma becsléseim szerint körülbelül százra tehető. Az egyes egységeken *belüli* huzalozás és a szinaptikus erősségek viszonylag stabilak és genetikailag meghatározottak - vagyis az egyes mintafelismerő modulokon belüli szerveződést genetikai tervezés határozza meg. A tanulás az ezen egységek *közötti* kapcsolatok létrehozásában történik, nem pedig azokon belül, és valószínűleg ezen egységek közötti kapcsolatok szinaptikus erősségében.

A tanulás alapmoduljának több tucat neuronból álló modulját a közelmúltban Henry Markram (1962-ben született) svájci idegtudós támogatta, akinek a teljes emberi agy szimulálására irányuló ambiciózus Blue Brain Projectjét a [7. fejezetben](#) ismertetem. Egy 2011-es tanulmányában leírja, hogy miközben a tényleges emlős neokortex neuronjait szkennelte és elemezte, "a kéreg legelemibb szintjén kereste a Hebb-összeállások bizonyítékát". Ehelyett - írja - "megfoghatatlan együtteseket talált, amelyeknek a kapcsolódása és a szinaptikus súlyok erősen kiszámíthatóak és korlátozottak". Arra a következtetésre jut, hogy "ezek az eredmények arra utalnak, hogy a tapasztalat nem tudja könnyen alakítani ezeket az összeállításoknak a szinaptikus kapcsolatait", és azt feltételezi, hogy "ezek a tudás veleszületett, Lego-szerű építőköveiként szolgálnak az érzékeléshez, és

hogy az emlékek elsajátítása ezeknek az építőköveknek a komplex konstrukciókká történő kombinálásával történik". Folytatja:

Funkcionális neuronális csoportosulásokról már évtizedek óta beszámoltak, de a szinaptikusan összekapcsolt neuronok klasztereinek közvetlen bizonyítékai... hiányoztak..... Mivel ezek a csoportok mind hasonló topológiájúak és szinaptikus súlyúak lesznek, és nem alakítja őket semmilyen konkrét tapasztalat, ezeket veleszületett csoportosulásoknak tekintjük..... A tapasztalat csak kisebb szerepet játszik a szinaptikus kapcsolatok és súlyok meghatározásában ezekben az elrendezésekben.... Tanulmányunk bizonyítékot talált [néhány tucat neuronból álló] veleszületett Lego-szerű összeállásokra.... Az egységek közötti összeköttetések szuper-összetételekben egyesíthetik őket egy neokortikális rétegen belül, majd magasabb rendű egységekben egy agykérgi oszlopban, még magasabb rendű egységekben egy agyi régióban, és végül a lehető legmagasabb rendű egységben, amelyet az egész agy képvisel.... Az emlékek elsajátítása nagyon hasonlít a Lego-építéshez. Minden egyes építmény egy-egy Lego-tömbnek felel meg, amely a világ feldolgozására, érzékelésére és a világra való reagálásra vonatkozó elemi, veleszületett tudás egy darabját tartalmazza.... Amikor különböző blokkok összeállnak, akkor ezeknek a veleszületett percepcióknak egy egyedi kombinációját alkotják, amely egy az egyén különleges tudása és tapasztalata.³

A Markram által javasolt "Lego blokkok" teljes mértékben megfelelnek az általam leírt mintafelismerő moduloknak. Egy e- mail kommunikációban Markram ezeket a "Lego blokkokat" "megosztott tartalomként és veleszületett tudásként" írta le. " ⁴Én úgy fogalmaznám meg, hogy ezeknek a moduloknak a célja a minták felismerése, megjegyzése és előrejelzése részleges minták alapján. Vegyük észre, hogy Markram becslése, miszerint az egyes modulok "több tucat neuront" tartalmaznak, csak a neokortex V. rétegén alapul. Az V. réteg valóban neuronokban gazdag, de a hat rétegben lévő neuronok számának szokásos arányát alapul véve ez modulonként körülbelül 100 neuron nagyságrendet jelentene, ami összhangban van az én becsléseimmel.

A neokortex konzisztens huzalozását és látszólagos modularitását már évek óta megfigyelték, de ez a tanulmány az első, amely bizonyítja e modulok stabilitását, miközben az agy dinamikus folyamatokon megy keresztül.

Egy másik, a Massachusetts General Hospital által végzett, a National Institutes of Health és a National Science Foundation által finanszírozott és a *Science* folyóirat 2012. márciusi számában megjelent tanulmány szintén a neokortexben lévő kapcsolatok szabályos szerkezetét mutatja.⁵ A cikk a következőképpen írja le a neokortex kábelezését rácsos mintázatot követve, mint a rendezett városi utcák: "Alapvetően az agy általános szerkezete Manhattanre hasonlít, ahol van egy kétdimenziós utcaterv és egy harmadik tengely, egy lift a harmadik dimenzióban" - írta Van J. Wedeen, a Harvard idegkutatója és fizikusa, a tanulmány vezetője.

A *Science* magazin podcastjában Wedeen ismertette a kutatás jelentőségét: "Ez az agyi pályák háromdimenziós szerkezetének vizsgálata volt. Amikor a tudósok az elmúlt

körülbelül száz évben az agy pályáiról gondolkodtak, az a tipikus kép vagy modell jutott eszükbe, hogy ezek a pályák egy tál spagettire hasonlítanak - különálló pályák, amelyek egymáshoz viszonyítva nem rendelkeznek különösebb térbeli mintázattal. Mágneses rezonanciás képalkotás segítségével ezt a kérdést kísérleti úton tudtuk megvizsgálni. És azt találtuk, hogy ahelyett, hogy véletlenszerűen elrendezett vagy független pályákról lenne szó, azt találtuk, hogy az agy összes pályája együttvéve egyetlen rendkívül egyszerű struktúrába illeszkedik. Alapvetően úgy néznek ki, mint egy kocka. Alapvetően három egymásra merőleges irányban futnak, és mindhárom irányban a pályák egymással nagyon párhuzamosak és tömbökbe rendeződnek. Tehát a független spagettik helyett azt látjuk, hogy az agy összekapcsolhatósága bizonyos értelemben egyetlen összefüggő struktúra."

Míg a Markram-tanulmány a neokortexben ismétlődő neuronmodulokat mutat, addig a Wedeen-tanulmány a modulok közötti kapcsolatok figyelemre méltóan rendezett mintázatát mutatja. Az agy kezdetben igen nagyszámú "várakozó kapcsolattal" rendelkezik, amelyekhez a mintafelismerő modulok csatlakozhatnak. Így ha egy adott modul egy másikhoz kíván kapcsolódni, nem kell axont növesztenie az egyikből és dendritet a másiktól, hogy a köztük lévő teljes fizikai távolságot áthidalja. Egyszerűen felhasználhatja az egyik ilyen várakozó axonális kapcsolatot, és egyszerűen csak rácsatlakozhat a szál végére. Ahogy Wedeen és kollégái írják: "Az agy pályái egy olyan alaptervet követnek, amelyet... a korai embriogenezis hozott létre. Így az érett agy pályái ennek a három őseredeti gradiensnek a képét mutatják, a fejlődés által fizikailag deformálva". Más szóval, ahogy tanulunk és tapasztalatokat szerzünk, a neokortex mintafelismerő moduljai ezekhez az előre kialakított kapcsolatokhoz kapcsolódnak, amelyek embrió korunkban jöttek létre.

Létezik egy hasonló elven alapuló elektronikus chip-típus, az úgynevezett FPGA (Field Programmable Gate Array). A chip több millió modult tartalmaz, amelyek logikai funkciókat valósítanak meg a várakozó kapcsolatokkal együtt. A használat idején ezeket a kapcsolatokat (elektronikus jelek segítségével) aktiválják vagy deaktiválják egy adott képesség megvalósítása érdekében.

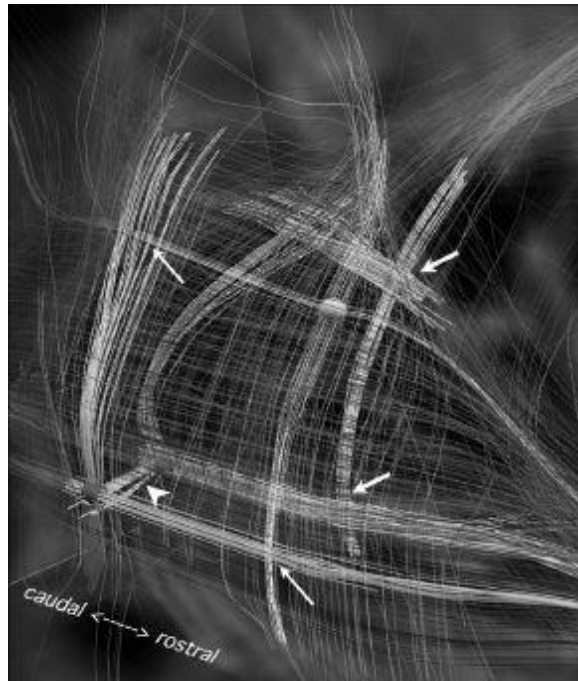
A neokortexben azok a hosszú távú kapcsolatok, amelyeket nem használnak, idővel eltűnnek, ami az egyik oka annak, hogy a neokortex egy közeli régiójának adaptálása a károsodott régió kompenzálására nem olyan hatékony, mint az eredeti régió használata. A Wedeen tanulmánya szerint a kezdeti kapcsolatok rendkívül rendezettek és ismétlődőek, akárcsak maguk a modulok, és rácsmintázatuk "irányítja a kapcsolódást" a neokortexben. Ezt a mintázatot az összes vizsgált főemlős és emberi agyban megtalálták, és az egész neokortexben megfigyelhető volt, a korai érzékszervi mintákkal foglalkozó régióktól egészen a magasabb szintű érzelmekig. Wedeen a *Science* folyóiratban megjelent cikkében arra a következtetésre jutott, hogy "az agyi pályák rácsszerkezete átható, összefüggő és folyamatos volt a fejlődés három fő tengelyével". Ez ismét egy közös algoritmusról beszél az összes neokortikális funkcióban.

Régóta ismert, hogy a neokortex legalábbis bizonyos régiói hierarchikusak. A legjobban tanulmányozott régió a látókéreg, amely V1, V2 és MT (más néven V5) területekre oszlik. Ahogy haladunk előre a magasabb területek felé ebben a régióban ("magasabb" a fogalmi feldolgozás értelmében, nem fizikailag, mivel a neokortex mindig csak egy mintafelismerő vastagságú), a felismerhető tulajdonságok egyre absztraktabbá

válnak. A V1 nagyon egyszerű éleket és primitív formákat ismer fel. A V2 felismeri a kontúrokat, az egyes szemek által bemutatott képek eltérését, a térbeli tájolást, és azt, hogy a kép egy része része része-e a képnek vagy sem.

egy tárgy vagy a háttér.⁶ A neokortex magasabb szintű régiói olyan fogalmakat ismernek fel, mint a tárgyak és arcok azonossága és mozgása. Az is régóta ismert, hogy a kommunikáció ezen a hierarchián keresztül felfelé és lefelé egyaránt zajlik, és hogy a jelek lehetnek gerjesztő és gátló jellegűek is. Tomaso Poggio (született 1947-ben), az MIT idegtudósa kiterjedten tanulmányozta az emberi agy látását, és az elmúlt harmincöt év kutatásai során

éveken keresztül jelentős szerepet játszott a hierarchikus tanulás és mintafelismerés kialakításában a vizuális neokortex "korai" (legalacsonyabb fogalmi) szintjein.⁷



A National Institutes of Health tanulmánya szerint a neokortexben a kezdeti kapcsolatok rendkívül szabályos rácsszerkezete.



A neokortikális kapcsolatok szabályos rácsszerkezetének egy másik nézete.



A neokortexben található rácsszerkezet figyelemre méltóan hasonlít az úgynevezett keresztsávós kapcsoláshoz, amelyet integrált áramkörökben és áramköri lapokban használnak.

A vizuális neokortex alacsonyabb hierarchikus szintjeiről alkotott képünk összhangban van az [előző fejezetben](#) ismertetett PRTM-mel, és a neokortex-feldolgozás hierarchikus természetének megfigyelése nemrégiben messze túlmutatott ezeken a szinteken. Daniel J. Felleman, a Texasi Egyetem neurobiológus professzora és munkatársai nyomon követték az agykéreg "hierarchikus szerveződését... [25 neokortikális területen]", amelyek között vizuális területek és magasabb szintű, több érzékszervből származó mintákat kombináló területek is voltak. Azt találták, hogy a neokortikális hierarchiában felfelé haladva a minták feldolgozása egyre absztraktabbá válik, nagyobb térbeli területeket foglal magában, és hosszabb időintervallumokat foglal magában. Minden egyes kapcsolatnál kommunikációt találtak mind felfelé, mind lefelé.

hierarchia.⁸

A legújabb kutatások lehetővé teszik, hogy ezeket a megfigyeléseket jelentősen kiterjesszük a látókérgen jóval túlmutató régiókra, sőt az asszociációs területekre is, amelyek több érzékszervből származó inputokat kombinálnak. Uri Hasson, a Princeton pszichológiaprofesszorának és munkatársainak 2008-ban közzétett tanulmánya azt mutatja, hogy a látókérgben megfigyelt jelenségek a neokortikális területek széles skáláján előfordulnak: "Jól ismert, hogy a vizuális agykérgi pályák mentén elhelyezkedő neuronok egyre nagyobb térbeli receptív mezőkkel rendelkeznek. Ez a vizuális rendszer egyik alapvető szervező elve.... A valós világ eseményei nemcsak a tér kiterjedt régióiban, hanem hosszabb időintervallumokban is előfordulnak. Ezért feltételeztük, hogy a térbeli receptív mezők

méretére talált hierarchiával analóg hierarchiának kell léteznie a különböző agyi régiók időbeli válaszcsoportjai esetében is." Pontosan ezt találták, ami lehetővé tette számukra azt a következtetést, hogy "a térbeli receptív mezők ismert agykérgi hierarchiájához hasonlóan az emberi agyban is létezik egy hierarchia a fokozatosan hosszabb időbeli receptív ablakok között.

agy. "9

A legerősebb érv a neokortexben történő feldolgozás egyetemessége mellett a plaszticitás (nem csak a tanulás, hanem a felcserélhetőség) mindenütt jelenlévő bizonyítékai. Más szóval, az egyik régió képes más régiók munkáját elvégezni, ami az egész neokortexben közös algoritmust feltételez. Az idegtudományi kutatások nagy része arra irányult, hogy meghatározzák, a neokortex mely régiói milyen típusú mintákért felelősek. Ennek meghatározására a klasszikus technika az volt, hogy kihatározzák a sérülésből vagy agyvérzésből származó agykárosodást, és az elvesztett funkciókat korrelálják az egyes sérült régiókkal. Így például, ha azt vesszük észre, hogy valaki, akinek a fusiform gyrus régiója újonnan szerzett károsodást, hirtelen nehezen ismeri fel az arcokat, de még mindig képes azonosítani az embereket a hangjuk és a nyelvi mintáik alapján, akkor feltételezhetjük, hogy ennek a régióknak köze van az arcfelismeréshez. Az alapfeltevés az volt, hogy minden egyes ilyen régiót egy bizonyos típusú minta felismerésére és feldolgozására tervezték. Bizonyos fizikai régiók bizonyos típusú mintákhoz társultak, mert normális körülmények között az információ így áramlik. Ha azonban ez a normális információáramlás bármilyen okból megszakad, a neokortex egy másik régiója képes átvenni a helyét.

A plaszticitást széles körben megfigyelték a neurológusok, akik megfigyelték, hogy a sérülés vagy stroke következtében agykárosodott betegek képesek újra megtanulni ugyanazokat a készségeket a neokortex egy másik területén. A plaszticitás talán legdrámaibb példája Marina Bedny amerikai idegtudós és munkatársai 2011-es tanulmánya arról, hogy mi történik a veleszületetten vak emberek látókéregével. Az általános bölcsesség szerint a látókéreg korai rétegei, mint például a V1 és V2, eredendően nagyon alacsony szintű mintákkal (például élekkel és görbékkel) foglalkoznak, míg a frontális kéreg (a kéregnek az az evolúciósan új régiója, amely egyedülállóan nagy homlokunkon található) eredendően a nyelv és más absztrakt fogalmak sokkal összetettebb és finomabb mintáival foglalkozik. De ahogy Bedny és kollégái megállapították: "Az embereknek feltehetően olyan agyi régiói fejlődtek ki a bal homlok- és halántékkéregben, amelyek egyedülállóan alkalmasak a nyelvi feldolgozásra. A veleszületetten vak egyének azonban bizonyos verbális feladatokban a látókérget is aktiválják. Bizonyítékot szolgáltatunk arra, hogy ez a vizuális kéreg aktivitás valójában a nyelvi feldolgozást tükrözi. Azt találtuk, hogy veleszületetten vak egyéneknél a bal oldali látókéreg a klasszikus nyelvi régiókhoz hasonlóan viselkedik.... Arra a következtetésre jutottunk, hogy olyan agyi régiók, amelyekről úgy gondoljuk, hogy a látáshoz fejlődtek ki, a korai tapasztalatok eredményeként átvehetik a nyelvi feldolgozást."10

Vegyük figyelembe a tanulmány következményeit: Ez azt jelenti, hogy fizikailag viszonylag távol eső neokortikális régiók, amelyeket fogalmilag is nagyon különbözőnek tartottak (primitív vizuális jelek kontra absztrakt nyelvi fogalmak), lényegében ugyanazt az algoritmust használják. Az ilyen eltérő típusú mintákat feldolgozó régiók helyettesíthetik

egymást.

Daniel E. Feldman, a Berkeley-i Kaliforniai Egyetem idegtudósa 2009-ben átfogó áttekintést írt az általa "a neokortex plaszticitásának szinaptikus mechanizmusai"-nak nevezett folyamatokról, és bizonyítékot talált az ilyen típusú plaszticitásra az egész neokortexben. Azt írja, hogy "a plaszticitás lehetővé teszi az agy számára, hogy megtanulja és megjegyezze az érzékszervi világ mintáit, hogy finomítsa a mozgásokat... és hogy sérülések után helyreállítsa a funkciókat". Hozzáteszi, hogy ezt a plaszticitást "a strukturális változások teszik lehetővé, beleértve a következők kialakulását, eltávolítását és morfológiai átalakulását

agykérgi szinapszisok és dendritikus gerincek".¹¹

A neokortikális plaszticitás (és ezáltal a neokortikális algoritmus egységességének) egy másik megdöbbentő példáját nemrégiben mutatták be a Berkeley-i Kaliforniai Egyetem tudósai. Beültetett mikroelektrodasorokat kapcsoltak be, hogy kimondottan az egerek motoros kéregének azon régiójából származó agyi jeleket vegyék fel, amely a bajszuk mozgását irányítja. Kísérletüket úgy állították be, hogy az egerek jutalmat kaptak, ha ezeket az idegsejteket úgy irányították, hogy egy bizonyos mentális mintában tüzeljenek, de ne mozdítsák meg ténylegesen a bajszukat. A jutalom elnyeréséhez szükséges minta olyan mentális feladatot jelentett, amelyet a frontális neuronjaik normális esetben nem végeznének el. Az egerek ennek ellenére képesek voltak ezt a mentális teljesítményt végrehajtani, lényegében úgy, hogy a motoros neuronjaikkal gondolkodtak, miközben mentálisan szétválasztották őket.

a motoros mozgások irányításától.¹² A következtetés az, hogy a motoros kéreg, a neokortexnek az izommozgások koordinálásáért felelős régiója is a standard neokortexi algoritmust használja.

Több oka is van azonban annak, hogy egy készség vagy egy tudásterület, amelyet a neokortex egy új területének felhasználásával tanultunk meg újra a sérült terület helyett, nem feltétlenül lesz olyan jó, mint az eredeti. Először is, mivel egy adott készség elsajátítása és tökéletesítése egy egész életen át tartott, az újratanulás a neokortex egy másik területén nem fog azonnal ugyanolyan eredményeket produkálni. Ennél is fontosabb, hogy a neokortexnek ez az új területe nem csak ül és várakozik a sérült terület helyett. Ez is létfontosságú funkciókat látott el, és ezért vonakodni fog feladni neokortikális mintáit, hogy a sérült régiót kompenzálja. Elkezdheti azzal, hogy felszabadítja mintáinak néhány felesleges másolatát, de ez finoman rontja a meglévő készségeit, és nem szabadít fel annyi agykérgi helyet, amennyit az újratanulandó készségek eredetileg használtak.

Van egy harmadik oka is annak, hogy a plaszticitásnak megvannak a maga korlátai. Mivel a legtöbb embernél bizonyos típusú minták meghatározott régiókon keresztül áramlanak (például az arcok feldolgozása a fusiform gyruson keresztül történik), ezek a régiók (a biológiai evolúció révén) optimalizálódtak az ilyen típusú mintákra. Amint arról a [7. fejezetben](#) beszámolok, ugyanezt az eredményt találtuk a digitális neokortikális fejlesztéseinkben is. A beszédet fel tudtuk ismerni a karakterfelismerő rendszereinkkel és fordítva, de a beszédrendszerek a beszédre voltak optimalizálva, és hasonlóképpen a karakterfelismerő rendszerek a nyomtatott karakterekre, így némi teljesítménycsökkenés

következtet be, ha az egyiket a másikkal helyettesítettük. Valójában evolúciós (genetikai) algoritmusokat használtunk ennek az optimalizálásnak a megvalósításához, ami a biológia természetes működésének szimulációja. Tekintettel arra, hogy az arcok a legtöbb embernél már több százezer éve (vagy még annál is több) átáramlanak a fusiform gyruson, a biológiai evolúciónak volt ideje arra, hogy kedvező képességet alakítson ki az ilyen minták feldolgozására ebben a régióban. Ez ugyanazt az alapalgoritmust használja, de az arcok felé orientálódik. Ahogy Randal Koene holland idegtudós írta: "A [neo]kéreg nagyon egységes, minden egyes oszlop vagy

minoszlop elvileg képes arra, amire a többi is." ¹³ Jelentős legújabb kutatások alátámasztják azt a megfigyelést, hogy a mintafelismerő modulok a rájuk ható minták alapján kábelezik be magukat. Yi Zuo idegtudós és munkatársai például megfigyelték, ahogyan az egerek egy új készség elsajátításakor (egy résen keresztül nyúlva magot ragadnak) új "dendritikus tüskék" alakultak ki az idegsejtek között. ¹⁴ A Salk Intézet kutatói felfedezték, hogy a neokortex moduljainak ez a kritikus önhuzamosítása a jelek szerint csak néhány gén irányítja. Ezek a gének és az önvezetékezésnek ez a módja is egységes az egész neokortexben. ¹⁵

A neokortex ezen tulajdonságait számos más tanulmány is dokumentálja, de foglaljuk össze, amit az idegtudományi szakirodalomból és a saját gondolat kísérleteinkből megfigyelhetünk. A neokortex alapegysége a neuronok modulja, amelyet én száz körüli becslök. Ezek minden egyes neokortikális oszlopban úgy vannak összeszövődve, hogy az egyes modulok nem különülnek el láthatóan egymástól. A kapcsolatok és a szinaptikus erősségek mintázata az egyes modulokon belül viszonylag stabil. A modulok közötti kapcsolatok és szinaptikus erősségek jelentik a tanulást.

A neokortexben mintegy kvadrillió (10^{15}) kapcsolat van, a genomban azonban csak körülbelül 25 millió bájtnyi tervezési információ található (veszteségmentes tömörítés után), ¹⁶ így maguk a kapcsolatok nem lehetnek genetikailag előre meghatározottak. Lehetséges, hogy a tanulás egy része a neokortex a régi agy kihallgatása, de ez még mindig szükségszerűen csak viszonylag kis mennyiségű információt jelentene. A modulok közötti kapcsolatok összességében a tapasztalatokból (inkább nevelésből, mint természetből) jönnek létre.

Az agy nem rendelkezik elegendő rugalmassággal ahhoz, hogy az egyes neokortikális mintafelismerő modulok egyszerűen összekapcsolódhassanak bármely más modullal (ahogyan azt a számítógépünkben vagy a weben könnyen be tudjuk programozni). -egy tényleges fizikai kapcsolatnak kell létrejönnie, amely egy dendrithez csatlakozó axonból áll. Mindannyian a lehetséges idegi kapcsolatok hatalmas készletével indulunk. Amint a Wedeen-tanulmány mutatja, ezek a kapcsolatok nagyon ismétlődő és rendezett módon szerveződnek. Az ezen axonokhoz való terminális kapcsolódás a várakozó axonokhoz azon minták alapján történik, amelyeket az egyes neokortikális mintafelismerők felismertek. A fel nem használt kapcsolatok végül elnyíródnak. Ezek a kapcsolatok hierarchikusan épülnek fel, tükrözve a valóság természetes hierarchikus rendjét. Ez a neokortex legfontosabb erőssége.

A neokortikális mintafelismerő modulok alapvető algoritmusa a neokortexben az "alacsony szintű" moduloktól, amelyek a legalapvetőbb érzékszervi mintákkal foglalkoznak, a "magas szintű" modulokig, amelyek a legabsztraktabb fogalmakat ismerik fel, egyenértékű. A neokortikális régiók plaszticitásának és felcserélhetőségének hatalmas bizonyítékai ezt a fontos megfigyelést igazolják. Van némi optimalizáció a bizonyos típusú mintákkal foglalkozó régiókban, de ez másodrendű hatás - az alapvető algoritmus univerzális. A jelek felfelé és lefelé haladnak a fogalmi hierarchiában. Egy felfelé menő jel azt jelenti, hogy "észleltem egy mintát". A lefelé menő jel azt jelenti, hogy "várom a mintád megjelenését", és lényegében egy előrejelzés. Mind a felfelé, mind a lefelé irányuló jelek lehetnek gerjesztő vagy gátló.

Minden minta egy meghatározott sorrendben van, és nem könnyen fordítható vissza. Még ha egy minta többdimenziósnak is tűnik, azt alacsonyabb szintű minták egydimenziós sorozata képviseli. Egy minta más minták rendezett sorozata, így minden felismerő eleve rekurzív. A hierarchiának sok szintje lehet.

Rengeteg redundancia van a megtanult mintákban, különösen a fontos mintákban. A minták (például a közös tárgyak és arcok) felismerése ugyanazt a mechanizmust használja, mint az emlékeink, amelyek csak megtanult minták. Ezek is minták sorozataként tárolódnak - ezek alapvetően történetek. Ezt a mechanizmust használjuk a tanuláshoz és a világban való fizikai mozgás végrehajtásához is. A minták redundanciája teszi lehetővé, hogy felismerjük a tárgyakat, embereket és gondolatokat, még akkor is, ha azok változatosak és különböző kontextusban fordulnak elő. A méret- és méretváltozékonysági paraméterek lehetővé teszik azt is, hogy a neokortex a különböző dimenziókkal (a hangok esetében az időtartam) szembeni nagyságbeli változásokat kódolja. Az egyik módja annak, hogy ezeket a nagyságparamétereket kódolni lehessen, egyszerűen több, különböző számú ismétlődő bemenetet tartalmazó mintázaton keresztül. Így például a "meredek" beszélt szóhoz létezhetnek olyan minták, amelyekben a hosszú magánhangzó [E] különböző számú ismétlése szerepel, mindegyikben a fontossági paramétert mérsékelt szintre állítva, ami azt jelzi, hogy az [E] ismétlése változó. Ez a megközelítés matematikailag nem egyenértékű az explicit méretparaméterekkel, és a gyakorlatban közel sem működik olyan jól, de a nagyság kódolásának egyik megközelítése. A legerősebb bizonyítékunk ezekre a paraméterekre az, hogy a mesterséges intelligencia rendszereinkben szükség van rájuk ahhoz, hogy az emberi szinthez közeli pontossági szinteket érjünk el.

A fenti összefoglaló képezi a következtetéseket, amelyeket a kutatási eredmények fentebb megosztott mintájából, valamint a korábban tárgyalt gondolat kísérletekből levont következtetésekből vonhatunk le. Állítom, hogy az általam bemutatott modell az egyetlen lehetséges modell, amely kielégíti a kutatás és a gondolat kísérleteink által megállapított összes korlátozást.

Végül, van még egy megerősítő bizonyíték. A mesterséges intelligencia területén az elmúlt évtizedek során kifejlesztett technikák, amelyek a valós világ jelenségeinek (például az emberi beszéd és az írott nyelv) felismerésére és intelligens feldolgozására, valamint a természetes nyelvű dokumentumok megértésére szolgálnak, matematikailag hasonlóan bizonyulnak a fent bemutatott modellhez. Ezek is a PRTM példái. A mesterséges

intelligencia területe nem kifejezetten az agy másolására törekedett, de ennek ellenére lényegében egyenértékű technikákhoz jutott.

FEJEZET 5 AZ ÖREG AGY

Az agyam öreg, de a memóriám kiváló.

-Al Lewis

Itt állunk ennek az új világnak a közepén, primitív agyunkkal, az egyszerű barlangi életre hangolódva, félelmetes erővel a rendelkezésünkre állva, amelyeket elég okosak vagyunk ahhoz, hogy felszabadítsunk, de amelyek következményeit nem tudjuk felfogni.

-Szent-Györgyi Albert

A régi agyunk - az, amelyik még emlősökké válásunk előtt volt - nem tűnt el. Valójában még mindig ez adja a motivációnk nagy részét a kielégülés keresésében és a veszélyek elkerülésében. Ezeket a célokat azonban a neokortexünk szabályozza, amely az emberi agyat mind tömegében, mind aktivitásában uralja.

Az állatok korábban neokortex nélkül is éltek és túléltek, sőt minden nem emlős állat ma is így tesz. Az emberi neokortexet tekinthetjük a nagy szublimátornak - így a nagy ragadozó elkerülésére irányuló primitív motivációnkat a neokortex ma átalakíthatja egy feladat elvégzésévé, hogy lenyűgözzük a főnökünket; a nagy vadászatból lehet egy könyv megírása, mondjuk, az elméről; a szaporodásra való törekvésből pedig lehet nyilvános elismerés megszerzése vagy a lakásunk díszítése. (Nos, ez utóbbi motiváció nem mindig

így elrejtve.)

A neokortex hasonlóképpen jól segít a problémamegoldásban, mivel képes pontosan modellezni a világot, ami tükrözi valódi hierarchikus természetét. De a régi agy az, amelyik ezeket a problémákat elénk tárja. Természetesen, mint minden okos bürokrácia, a neokortex is gyakran úgy kezeli a neki kijelölt problémákat, hogy újrafogalmazza azokat. Ennek kapcsán tekintsük át a régi agy információfeldolgozását.

Az érzékszervi útvonal

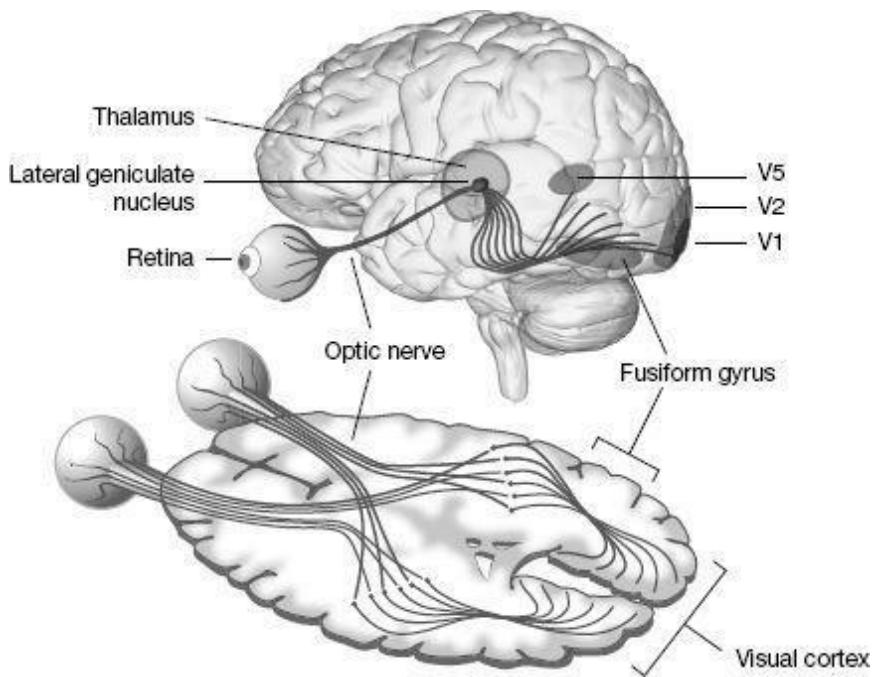
Az agyban lévő látóidegek rostjai mentén mozgással terjedő képek okozzák a látást.

-Isaac Newton

Mindannyian a saját agyunk univerzumában - börtönében - élünk. Ebből milliányi törékeny érzékelő idegszál nyúlik ki, melyek csoportjai egyedülállóan alkalmasak a körülöttünk lévő világ energetikai állapotainak - hő, fény, erő és kémiai összetétel - érzékelésére. Ez minden, amit közvetlenül tudunk róla; minden más csak logikus következtetés.

-Vernon Mountcastle¹

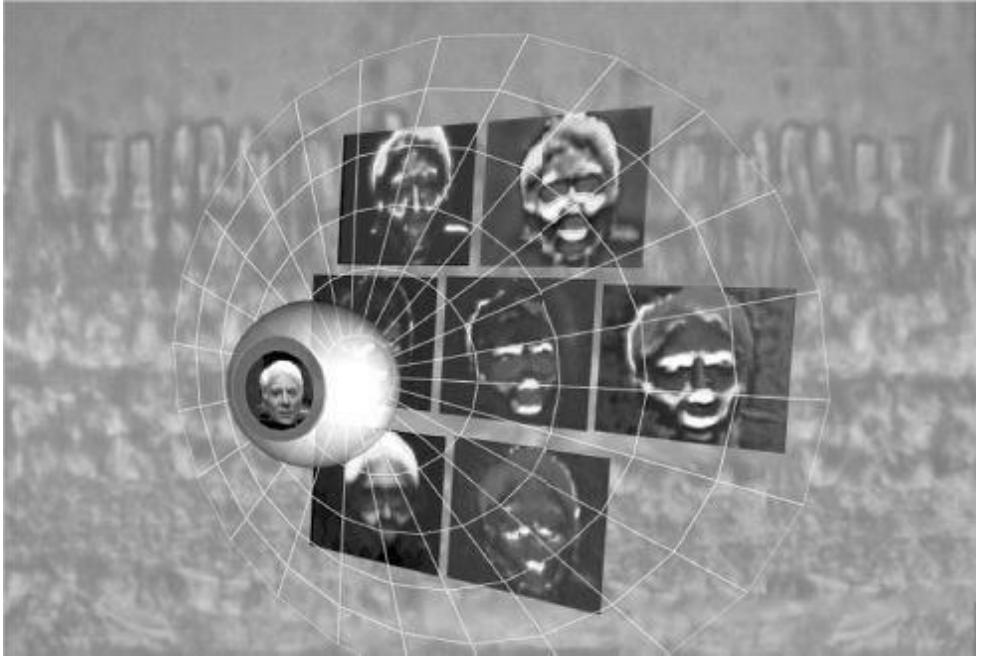
Bár azt az illúziót éljük át, mintha nagy felbontású képeket kapnánk a szemünkből, a látóideg valójában csak egy sor körvonalat és utalást küld az agynak a látómezőnkben lévő érdekes pontokról. Ezután lényegében hallucináljuk a világot az agykérgi memóriákból, amelyek párhuzamos csatornákon érkező, nagyon alacsony adatátviteli sebességű filmek sorozatát értelmezik. Frank S. Werblin, a berkeley-i Kaliforniai Egyetem molekuláris és sejtbiológia professzora és Dr. Roska Boton doktorandusz a *Nature* folyóiratban megjelent tanulmányában kimutatta, hogy a látóideg tíz-tizenkét kimeneti csatornát hordoz, amelyek mindegyike csak kis mennyiségű információt hordoz egy adott jelenetről.² Az úgynevezett ganglionsejtek egyik csoportja csak az élekről (kontrasztváltozásokról) küld információt. Egy másik csoport csak az egységes színű nagy területeket érzékeli, míg egy harmadik csoport csak az érdekes alakok mögötti háttérre érzékeny.



A vizuális pálya az agyban.

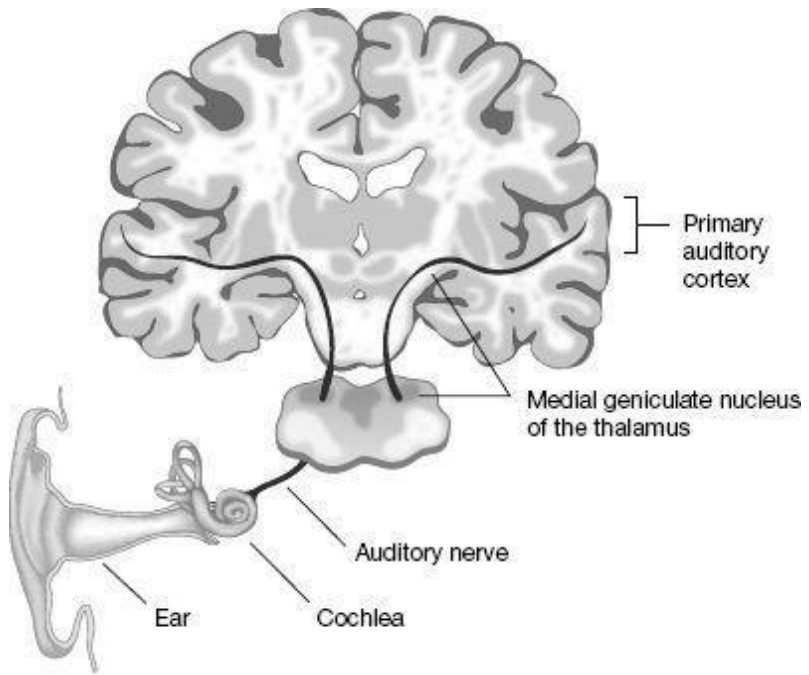
"Bár azt hisszük, hogy a világot teljes egészében látjuk, amit kapunk, valójában csak utalások, élek térben és időben" - mondja Werblin. "A világnak ez a 12 képe alkotja az összes információt, amivel valaha is rendelkezni fogunk arról, hogy mi van odakint, és ebből a 12 képből, amelyek olyan gyéren vannak, rekonstruáljuk a vizuális világ gazdagságát. Kíváncsi vagyok, hogyan választotta ki a természet ezt a 12 egyszerű filmet, és hogyan lehetséges, hogy ezek elegendőek ahhoz, hogy minden információval ellássanak minket, amire látszólag szükségünk van"."

Ezt az adatsökkentést a mesterséges intelligencia területén "ritka kódolásnak" nevezzük. A mesterséges rendszerek létrehozásakor azt tapasztaltuk, hogy a bemeneti információ nagy részének kidobása és csak a legmarkánsabb részletek megtartása kiváló eredményeket hoz. Ellenkező esetben a neokortex (biológiai vagy más) korlátozott információfeldolgozó képessége túlterhelődik.



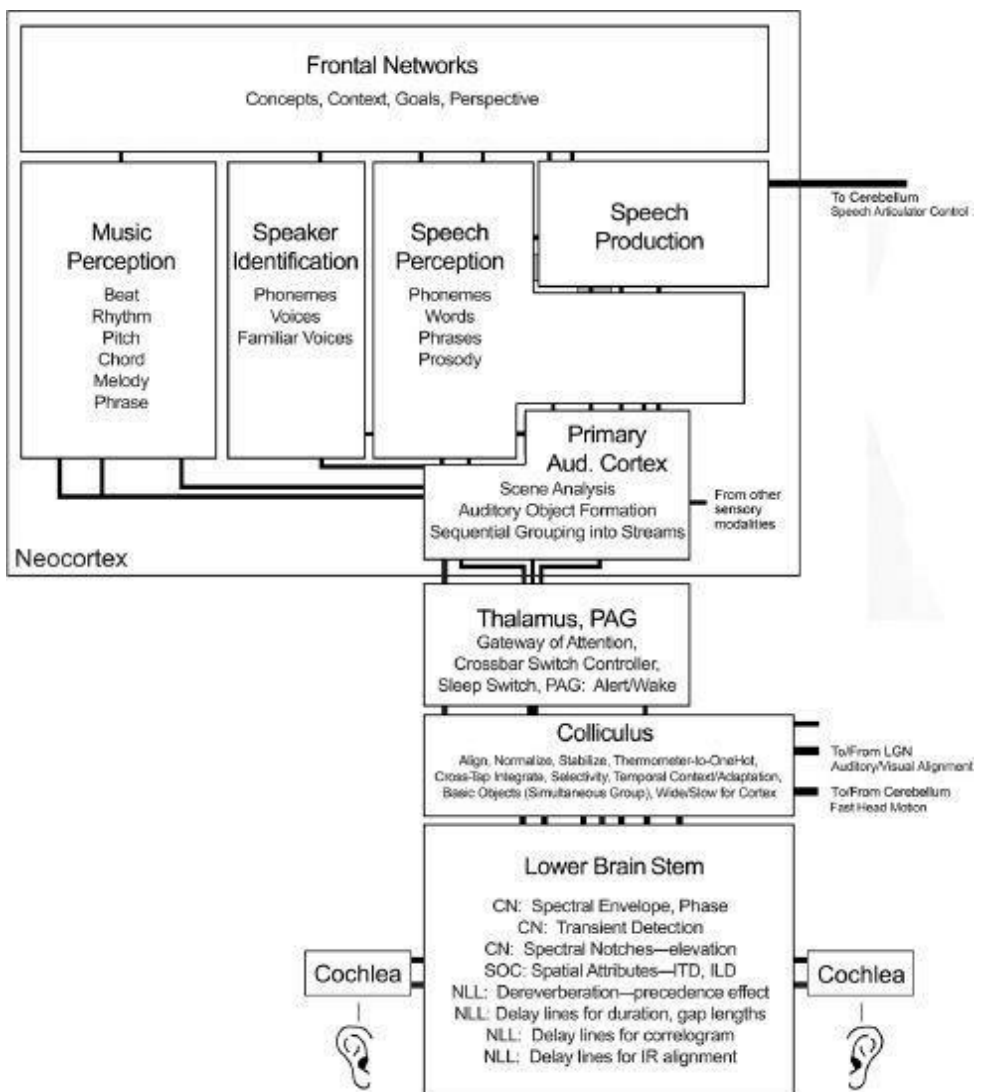
A tizenkét alacsony adatátviteli sebességű "film" közül hét, amelyet a látóideg küld az agyba.

Lloyd Watts és az Audience, Inc. kutatócsoportja aprólékosan modellezte a hallási információk feldolgozását az emberi csigahártyától a szubkortikális régiókon, majd a neokortex korai szakaszain keresztül.³ Olyan kutatási technológiát fejlesztettek ki, amely különböző 600 frekvenciasávokat (oktávonként 60) von ki a hangból. Ez sokkal közelebb áll az emberi csiga által kivont 3000 sávra becsült értékhez (szemben a kereskedelmi forgalomban kapható beszédfelismerővel, amely csak 16-32 sávot használ). Az Audience két mikrofón és a hallásfeldolgozás részletes (és nagy spektrális felbontású) modellje segítségével olyan kereskedelmi technológiát hozott létre (valamivel alacsonyabb spektrális felbontással, mint a kutatási rendszere), amely hatékonyan távolítja el a háttérzajt a beszélgetésekből. Ezt ma már számos népszerű mobiltelefonban használják, és lenyűgöző példája egy olyan kereskedelmi terméknek, amely annak megértésén alapul, hogy az emberi hallásérzékelő rendszer hogyan képes egyetlen érdekes hangforrásra összpontosítani.



A hallási pálya az agyban.

A testből érkező bemenetek (becslések szerint másodpercenként több száz megabit), beleértve a bőrből, izmokból, szervekből és más területekről érkező idegeket, a felső gerincvelőbe áramlanak. Ezek az üzenetek nem csupán az érintésről szóló kommunikációt foglalják magukban; emellett információt hordoznak a hőmérsékletről, a savszintről (például az izmok tejsavszintjéről), a tápláléknak a gyomor-bélrendszerben történő mozgásáról és sok más jelről. Ezeket az adatokat az agytörzs és a középagy dolgozza fel. A lamina 1 neuronoknak nevezett kulcssejtek létrehozzák a test térképét, amely a test aktuális állapotát ábrázolja, hasonlóan a repülésirányítók által a repülőgépek követésére használt kijelzőkhöz. Innen az érzékszervi adatok egy titokzatos régióba, a talamuszba jutnak, és ezzel el is érkeztünk a következő témánkhoz.



A hallásfeldolgozás egyszerűsített modellje mind a szubkortikális területeken (a neokortex előtti területek), mind a neokortexben, amelyet az Audience, Inc. Az ábra L. Watts, "Reverse-Engineering the Human Auditory Pathway" című könyvéből származik.

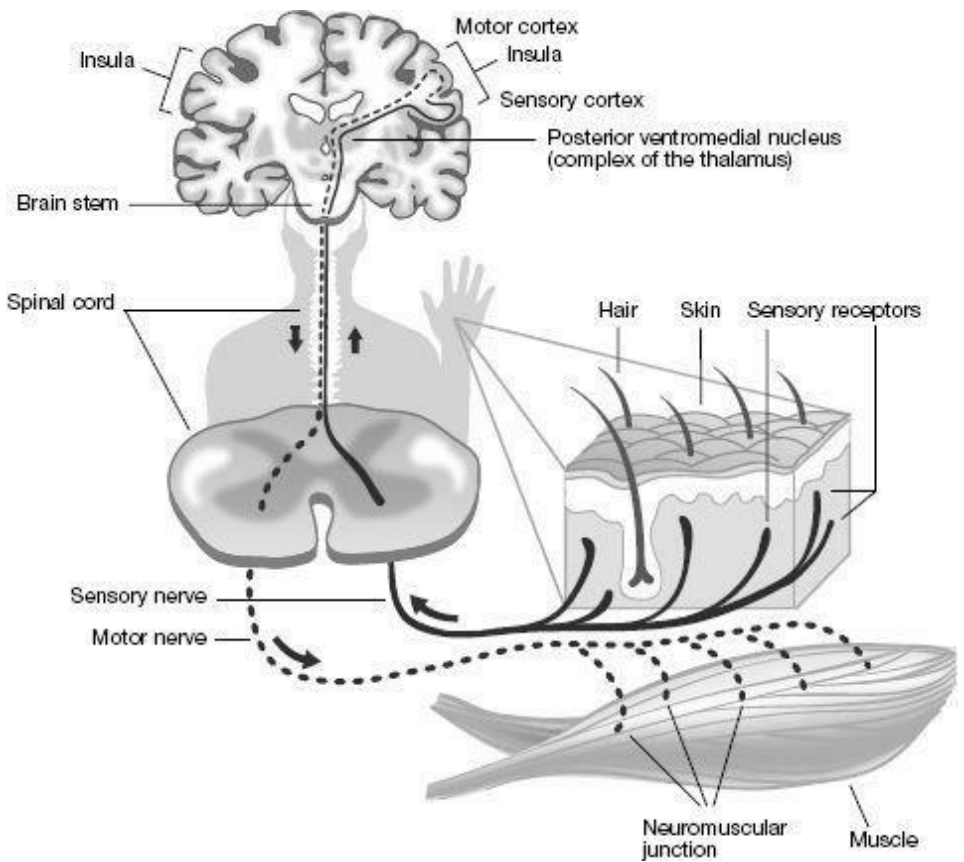
J. Liu et al. (szerk.), *WCCI 2012* (Berlin: Springer-Verlag, 2012), 49. o.

A thalamus

Mindenki tudja, mi a figyelem. Ez az elme által tiszta és élénk formában történő birtokba vétele egy, a látszólag egyszerre több lehetséges tárgy vagy gondolatmenet közül. A tudat fókuszálása, koncentrációja a lényege. Egyes dolgoktól való elvonulást jelent, hogy hatékonyan foglalkozhassunk másokkal.

-William James

A középgagyból az érzékszervi információk ezután a thalamus hátsó ventromedialis magjának (VMpo) nevezett diónyi területen keresztül áramlanak, amely a testi állapotokra adott komplex reakciókat számítja ki, mint például "ennek szörnyű íze van", "micsoda bűz" vagy "ez a könnyű érintés ingerlő". Az egyre inkább feldolgozott információ a neokortex két régiójába, az insulába kerül. Ezek a kisujjnyi méretű struktúrák a neokortex bal és jobb oldalán helyezkednek el. Dr. Arthur Craig, a phoenixi Barrow Neurológiai Intézet munkatársa a VMpo-t és a két insula régiót úgy írja le, mint "egy rendszert amely az anyagi éneket képviseli."⁴



Az érzékelés-érintés útvonala az agyban.

Egyéb funkciói mellett a talamuszt tekintik az előfeldolgozott érzékszervi információknak a neokortexbe való bejutásának kapujának. A VMpo-n keresztül áramló tapintási információn kívül a látóidegből származó feldolgozott információ (amely, mint fentebb említettük, már jelentősen átalakult) a talamusz laterális geniculáris magnak nevezett régiójába kerül, amely aztán továbbküldi azt a neokortex V1 régiójába. A hallásérzékelésből származó információ a talamusz medialis geniculáris magján keresztül jut el a neokortex korai hallási régióiba. Minden érzékszervi adatunk (kivéve, úgy tűnik, a szaglórészt, amely a szaglógumót használja helyette) a talamusz meghatározott régióin keresztül halad.

A talamusz legjelentősebb szerepe azonban a neokortexszel való folyamatos kommunikáció. A neokortexben lévő mintafelismerők előzetes eredményeket küldenek a talamusznak, és a válaszokat alapvetően az egyes felismerők VI. rétegéből érkező gerjesztő és gátló kölcsönös jelek segítségével kapják. Ne feledjük, hogy ezek nem vezeték nélküli üzenetek, így a neokortex és a talamusz valamennyi régiója között rendkívüli mennyiségű tényleges vezeték (axonok formájában) van szükség. Gondoljunk csak arra, hogy a neokortexben lévő több százmillió mintafelismerőnek milyen hatalmas területre van szüksége (a szükséges kapcsolatok fizikai tömegét tekintve) ahhoz, hogy folyamatosan jelentkezzenek a talamussal.⁵

Miről beszélgetnek tehát a neokortikális mintafelismerők százmilliói a talamussal? A jelek szerint fontos beszélgetésről, mert a talamusz fő régiójának kétoldali mélyreható károsodása elhúzódó eszméletvesztéshez vezethet. Egy sérült talamuszú embernek még mindig lehet aktivitása a neokortexében, azáltal, hogy az asszociációval történő önindító gondolkodás még működhet. De az irányított gondolkodás - az a fajta, ami miatt kikészülődünk az ágyból, beülünk az autónkba, és leülünk az íróasztalunkhoz a munkahelyünkön - nem működik talamusz nélkül. Egy híres esetben a huszonegy éves Karen Ann Quinlan szívrohamot és légzési elégtelenséget szenvedett, és tíz éven át nem reagáló, látszólag vegetatív állapotban maradt. Amikor meghalt, a boncolás során kiderült, hogy a neokortexe normális volt, de a talamusa elpusztult.

A thalamus a neokortexben található strukturált tudásra támaszkodik ahhoz, hogy betölthesse a figyelem irányításában betöltött kulcsszerepét. Képes egy (a neokortexben tárolt) listán végigmenni, lehetővé téve számunkra, hogy kövessünk egy gondolatmenetet vagy egy cselekvési tervet. Az MIT Picower Institute for Learning and Memory neurológusainak legújabb kutatásai szerint egyszerre körülbelül négy elemet tudunk a munkamemóriánkban tartani, féltekéntként kettőt.⁶ Az, hogy a talamusz irányítja-e a neokortexet vagy fordítva, korántsem egyértelmű, de mindkettő nélkül képtelenek vagyunk működni.

A Hippocampus

Minden agyfélteke tartalmaz egy hippocampuszt, egy kis régiót, amely úgy néz ki, mint egy tengeri ló, a középső halántéklebenyben. Elsődleges funkciója az új eseményekre való emlékezés. Mivel az érzékszervi információk a neokortexen keresztül áramlanak, a neokortex feladata annak megállapítása, hogy egy élmény újszerű-e, hogy azt a hippocampusz számára bemutassa. Ezt vagy úgy éri el, hogy nem ismeri fel a jellemzők egy bizonyos csoportját (például egy új arcot), vagy úgy, hogy rájön, hogy egy egyébként ismerős helyzet most egyedi tulajdonságokkal rendelkezik (például a házastársunk álbajuszt visel).

A hippocampusz képes emlékezni ezekre a helyzetekre, bár úgy tűnik, hogy ezt elsősorban a neokortexbe mutató mutatókon keresztül teszi. Tehát a hippocampuszban az emlékek alacsonyabb szintű mintázatokként is tárolódnak, amelyeket korábban a neokortexben ismertek fel és tároltak el. Az érzékszervi tapasztalatokat moduláló neokortexszel nem rendelkező állatok esetében a hippocampusz egyszerűen az érzékszervekből származó információkra emlékezik, bár ezek már átmentek érzékszervi előfeldolgozáson (például a látóideg által végzett átalakításokon).

Bár a hippocampusz a neokortextet használja (ha az adott agy rendelkezik ilyennel), mint íróasztalt, a memóriája (a neokortexbe mutató mutatók) nem eredendően hierarchikus. A neokortex nélküli állatok ennek megfelelően emlékezhetnek dolgokra a hippocampuszuk segítségével, de az ő emlékeik nem lesznek hierarchikusak.

A hippocampusz kapacitása korlátozott, ezért a memóriája rövid távú. A rövid távú memóriából egy adott mintasorozatot úgy visz át a neokortex hosszú távú hierarchikus memóriájába, hogy ezt az emléksorozatot újra és újra lejátssza a neokortexnek. Szükségünk van tehát a hippocampuszra ahhoz, hogy új emlékeket és készségeket tanuljunk (bár úgy tűnik, hogy a szigorúan vett motoros készségek más mechanizmust használnak). Valaki, akinek mindkét hippocampusz-kópiája károsodott, megtartja a meglévő emlékeit, de nem lesz képes újakat kialakítani.

Theodore Berger, a Dél-kaliforniai Egyetem idegkutatója és kollégái modellezték egy patkány hippocampuszát, és sikeresen kísérleteztek egy mesterséges hippocampusz beültetésével. Egy 2011-ben ismertetett tanulmányukban az USC tudósai gyógyszerekkel blokkoltak bizonyos tanult viselkedéseket patkányokban. A mesterséges hippocampusz segítségével a patkányok gyorsan képesek voltak újratanulni a viselkedést. "Kapcsoljuk be a kapcsolót, és a patkányok emlékeznek. Kapcsoljuk ki, és a patkányok elfelejtik" - írta Berger, utalva arra, hogy az idegi implantátumot távolról is képes volt irányítani. Egy másik kísérletben a tudósok a patkányok természetes hippocampuszával együtt működtették a mesterséges hippocampuszukat. Az eredmény az lett, hogy a patkányok képessége az új viselkedések megtanulására megerősödött. "Ezek az integrált kísérleti modellezési vizsgálatok először mutatják meg" - magyarázta Berger - "hogyan... egy olyan neurális protézis, amely képes a kódolási folyamat valós idejű azonosítására és manipulálására, helyreállíthatja, sőt fokozhatja a kognitív mnemonikus

folyamatok."² A hippocampusz az Alzheimer-kór által elsőként károsított régiók egyike,

ezért a kutatás egyik célja olyan neurális implantátum kifejlesztése az emberek számára, amely enyhíti a betegség okozta károsodásnak ezt az első fázisát.

A kisagy

Kétféleképpen lehet elkapni egy röpülő labdát. Megoldhatod a labda mozgását irányító bonyolult szimultán differenciálegyenleteket, valamint a labdára való rálátásod sajátos szögét szabályozó további egyenleteket, majd még több egyenletet számolhatsz ki arra vonatkozóan, hogyan mozgassd a tested, a karod és a kezed, hogy a megfelelő időben a megfelelő helyen legyél.

Az agyad nem ezt a megközelítést alkalmazza. Alapvetően leegyszerűsíti a problémát azáltal, hogy egy csomó egyenletet egy egyszerű trendmodellbe zsugorít, figyelembe véve annak trendjeit, hogy a labda hol látszik a látómeződben, és milyen gyorsan mozog azon belül. Ugyanezt teszi a kezeddal is, lényegében lineáris előrejelzéseket készít a labda látszólagos helyzetére a látómeződben és a kezედben. A cél természetesen az, hogy ezek a két pont térben és időben ugyanabban a pontban találkozzanak. Ha úgy tűnik, hogy a labda túl gyorsan esik, a kezéd pedig túl lassan mozog, az agyad arra fogja irányítani a kezéd, hogy gyorsabban mozogjon, hogy a tendenciák egybeessenek. Ezt a "gordiuszi csomó" megoldást az egyébként megoldhatatlan matematikai problémára alapfüggvényeknek nevezik, és a kisagy végzi őket, egy bab alakú és megfelelő módon baseball-labda méretű régió, amely az agytörzsön helyezkedik el.⁸

A kisagy egy olyan régi agyi régió, amely egykor gyakorlatilag az összes emberi mozgást irányította. Még mindig az agyban található neuronok felét tartalmazza, bár ezek többsége viszonylag kicsi, így a régió az agy tömegének csak mintegy 10 százalékát teszi ki. A kisagy szintén a masszív ismétlődés egy másik példáját képviseli az agy felépítésében. A genomban viszonylag kevés információ található a felépítéséről, mivel szerkezete több neuron milliárdszor ismétlődő mintázata. Ahogyan a neokortexben, annak szerkezetében egységesség van.⁹

Az izmaink irányításának nagy részét a neokortex vette át, ugyanazokat a mintafelismerő algoritmusokat használva, amelyeket az érzékeléshez és a megismeréshez is használ. A mozgás esetében a neokortex funkcióját megfelelőbben nevezhetnénk mintaimplementációnak. A neokortex valóban használja a kisagyban lévő memóriát a mozgások finom forgatókönyveinek rögzítéséhez - például az Ön aláírása és bizonyos díszítések a művészi kifejezésben, például a zenében és a táncban. A kisagy szerepét a gyermekek kézírás-tanulása során vizsgáló tanulmányok azt mutatják, hogy a kisagy Purkinje-sejtjei valójában a mozdulatsorozatot mintavételezik, a

mindegyik érzékeny egy adott mintára.¹⁰ Mivel mozgásunk nagy részét ma már a neokortex irányítja, sok ember még a kisagy jelentős károsodása esetén is viszonylag szerény nyilvánvaló fogyatékossgal tud boldogulni, kivéve, hogy mozgásaik kevésbé kecsesek lehetnek.

A neokortex a kisagyat is felkérheti arra, hogy használja a valós idejű alapfüggvények kiszámításának képességét, hogy előre jelezze, milyen eredményei lennének azoknak a cselekvéseknek, amelyeket fontolgatunk, de még nem hajtottunk végre (és talán soha nem is fogunk végrehajtani), valamint mások cselekvéseinek vagy lehetséges cselekvéseinek. Ez egy újabb példa az agyban veleszületett beépített lineáris előrejelzőkre.

Jelentős előrelépés történt a kisagy szimulációjában a tekintetben, hogy a kisagy dinamikusan reagáljon az érzékszervi jelzésekre a fent tárgyalt alapfüggvények segítségével, mind az alulról felfelé irányuló szimulációkban (biokémiai modelleken alapuló), mind a felülről lefelé irányuló szimulációkban (matematikai modelleken alapuló, hogy az egyes ismétlődő jelzések hogyan reagálnak a kisagyra).¹¹

Élvezet és félelem

A félelem a babona fő forrása, és a kegyetlenség egyik fő forrása. A félelem legyőzése a bölcsesség kezdete.

-Bertrand Russell

Érezd a félelmet, és tedd meg mégis.

-Susan Jeffers

Ha a neokortex jó a problémamegoldásban, akkor mi a fő probléma, amit meg akarunk oldani? A probléma, amelyet az evolúció mindig is megpróbált megoldani, a faj fennmaradása. Ez az egyén túlélését jelenti, és mindannyian a saját neokortexünket használjuk arra, hogy ezt számtalan módon értelmezzük. A túlélés érdekében az állatoknak meg kell szerezniük a következő ételt, ugyanakkor el kell kerülniük, hogy valaki más

ételeivé váljanak. Szaporodniuk is kell. A legkorábbi agyakban olyan öröm- és félelemrendszerek fejlődtek ki, amelyek jutalmazták ezeknek az alapvető szükségleteknek a kielégítését, valamint az ezeket elősegítő alapvető viselkedési formákat. Ahogy a környezet és a konkurens fajok fokozatosan megváltoztak, a biológiai evolúció is ennek megfelelő változásokat hajtott végre. A hierarchikus gondolkodás megjelenésével a kritikus késztetések kielégítése összetettebbé vált, mivel most már az eszméken belüli eszmék hatalmas komplexumának volt alárendelve. De a neokortex általi jelentős modulációja ellenére a régi agy még mindig él és virul, és még mindig motivál bennünket örömmel és félelemmel.

Az egyik régió, amely az élvezethez kapcsolódik, a nucleus accumbens. Az 1950-es években végzett híres kísérletek során a patkányok, amelyek képesek voltak közvetlenül stimulálni ezt a kis régiót (egy beültetett elektródákat aktiváló kar megnyomásával), ezt minden mással szemben előnyben részesítették, beleértve a szexet vagy az evést, végül kimerültek és halálra éhezették magukat.¹² Embereknél más régiók is részt vesznek az élvezetben, például a ventrális pallidum és természetesen maga a neokortex.

Az örömet olyan vegyi anyagok is szabályozzák, mint a dopamin és a szerotonin. E könyv kereteit meghaladja, hogy részletesen tárgyaljuk ezeket a rendszereket, de fontos felismerni, hogy ezeket a mechanizmusokat preemlős rokonainktól örököltük. A neokortexünk feladata, hogy képessé tegyen bennünket arra, hogy az öröm és a félelem urai legyünk, ne pedig rabszolgái. A neokortex nem mindig jár sikerrel ebben a törekvésében, amilyen mértékben gyakran függőségi viselkedésnek vagyunk kitéve. A dopamin különösen az öröm megélésében részt vevő neurotranszmitter. Ha bármi jó történik velünk megnyerjük a lottót, kivívjuk társaink elismerését, megölel minket egy szerettünk, vagy akár olyan finom eredmények, mint például egy barátunk nevetése egy viccen -, dopamin felszabadulást tapasztalunk. Néha mi is, akárcsak a patkányok, akik a nucleus accumbensüket túlstimulálva haltak meg, rövidített utat használunk az örömkítőresek eléréséhez, ami nem mindig jó ötlet.

A szerencsejáték például dopamint szabadíthat fel, legalábbis ha nyerünk, de ez a szerencsejátékban rejlő kiszámíthatóság hiányától függ. A szerencsejáték egy ideig működhet a dopamin felszabadítása céljából, de mivel az esélyek szándékosan ellened vannak beállítva (különben a kaszinók üzleti modellje nem működne), rendszeres stratégiaként tönkretelhet. Hasonló veszélyek bármely függőséget okozó viselkedéssel együtt járnak. A dopaminreceptor D2 gén egy bizonyos genetikai mutációja különösen erős örömrészt okoz a függőséget okozó anyagokkal és viselkedésekkel kapcsolatos kezdeti tapasztalatokból, de mint köztudott (de nem mindig figyelnek oda eléggé), ezeknek az anyagoknak a későbbi használat során fokozatosan csökken az örömrészt kiváltó képessége. Egy másik genetikai mutáció azt eredményezi, hogy az emberek nem kapnak normális szintű dopaminfelszabadulást a mindennapi teljesítményekből, ami szintén a függőséget okozó tevékenységekkel kapcsolatos fokozott korai tapasztalatok kereséséhez vezethet. A népesség azon kisebbsége, amely ilyen genetikai hajlamot mutat a függőségre, hatalmas társadalmi és orvosi problémát okoz. Még azok is, akiknek sikerül elkerülniük a súlyosan addiktív viselkedést, küzdenek azzal, hogy egyensúlyt teremtsenek a dopaminfelszabadulás jutalmainak és az azt felszabadító viselkedés következményeinek

között.

A szerotonin egy olyan neurotranszmitter, amely fontos szerepet játszik a hangulat szabályozásában. Magasabb szintjén a jólét és az elégedettség érzésével jár együtt. A szerotonnának más funkciói is vannak, többek között a szinaptikus erősség, az étvágy, az alvás, a szexuális vágy és az emésztés modulálása. Az antidepresszív szereknek, például a szelektív szerotonin-visszavétel-gátlóknak (amelyek általában növelik a receptorok számára elérhető szerotoninszintet) általában messzemenő hatásai vannak, amelyek közül nem mindegyik kívánatos (például a libidó elnyomása). A neokortexben kifejtett hatásokkal ellentétben, ahol a minták felismerése és az axonok aktiválása egyszerre csak néhány neokortikális áramkörre hat, ezek az anyagok az agy nagy régióira vagy akár az egész idegrendszerre hatnak.

Az emberi agy mindkét féltekéjén található egy-egy amigdala, amely egy mandula alakú, több kis lebenyből álló régióból áll. Az amigdala szintén az öregagy része, és számos típusú érzelmi reakció feldolgozásában vesz részt, amelyek közül a legjelentősebb a félelem. Az ősemlős állatokban bizonyos előre programozott, veszélyt jelentő ingerek közvetlenül az amygdalába táplálkoznak, amely viszont beindítja a "harcolj vagy menekülj" mechanizmust. Az emberben az amigdala ma már a neokortex által közvetített veszélyérzetektől függ. A főnöke negatív megjegyzése például kiválthat egy ilyen választ azáltal, hogy a munkahelye elvesztésétől való félelmet generálja (vagy talán nem, ha bízunk a B-tervben). Amint az amigdala valóban úgy dönt, hogy veszély fenyeget, egy ősi eseménysorozat következik be. Az amigdala jelzi az agyalapi mirigynek, hogy szabadítson fel egy ACTH (adrenokortikotropin) nevű hormont. Ez viszont a mellékveséből a kortizol nevű stresszhormont váltja ki, ami azt eredményezi, hogy több energiát kapnak az izmok és az idegrendszer. A mellékvesék adrenalint és noradrenalint is termelnek, amelyek elnyomják az emésztő-, immun- és szaporodási rendszeredet (gondolván, hogy ezek vészhelyzetben nem kiemelt fontosságú folyamatok). A vérnyomás, a vércukor, a koleszterin és a fibrinogén (amely felgyorsítja a véralvadást) szintje emelkedik. A szívritmus és a légzés felgyorsul. Még a pupillái is kitágulnak, így jobban látja az ellenséget vagy a menekülési útvonalat. Mindez nagyon hasznos, ha hirtelen valós veszély, például egy ragadozó keresztezi az utadat. Köztudott, hogy a mai világban ennek a harcolj vagy menekülj mechanizmusnak a krónikus aktiválása maradandó egészségkárosodáshoz vezethet a magas vérnyomás, a magas koleszterinszint és egyéb problémák tekintetében.

A globális neurotranszmitterszintek, mint például a szerotonin, és a hormonszintek, mint például a dopamin, rendszere bonyolult, és a könyv hátralévő részét ezzel a témával tölthetnénk (ahogyan azt már nagyon sok könyv tette), de érdemes rámutatni, hogy az információ sávszélessége (az információfeldolgozás sebessége) ebben a rendszerben nagyon alacsony a neokortex sávszélességéhez képest. Csak korlátozott számú anyag vesz részt benne, és ezeknek a vegyi anyagoknak a szintje általában lassan változik, és viszonylag univerzális az egész agyban, szemben a neokortexszel, amely több száz trillió kapcsolatból áll, amelyek gyorsan változhatnak.

Joggal mondhatjuk, hogy érzelmi élményeink a régi és az új agyunkban egyaránt zajlanak. A gondolkodás az új agyban (a neokortexben) zajlik, de az érzés mindkettőben. Az emberi viselkedés bármilyen utánzásának ezért mindkettőt modelleznie kell. Ha azonban

csak az emberi kognitív intelligenciát keressük, a neokortex elegendő. A régi agyat helyettesíthetjük a nem biológiai neokortex közvetlenebb motivációjával, hogy elérjük az általunk kijelölt célokat. Watson esetében például a célt egyszerűen fogalmaztuk meg: Jöjjen ki helyes válaszokkal a *Jeopardy!* kérdésekre (bár ezeket tovább modulálta egy olyan program, amely megértette a *Jeopardy!* fogadásokat). A Nuance és az IBM által közösen fejlesztett új Watson-rendszer esetében az orvosi ismeretekhez a cél az emberi betegségek kezelésének segítése. A jövőbeli rendszereknek olyan céljaik lehetnek, mint a betegségek tényleges gyógyítása és a szegénység enyhítése. Az öröm és félelem harcának nagy része már elavult az ember számára, mivel a régi agy már jóval azelőtt kifejlődött, hogy még a primitív emberi társadalom is elindult volna; valójában a legtöbbje hullószerű.

Az emberi agyban folyamatos küzdelem folyik arról, hogy a régi vagy az új agy irányítson. A régi agy próbálja meghatározni a napirendet az öröm- és félelemélmények irányításával, míg az új agy folyamatosan próbálja megérteni a régi agy viszonylag primitív algoritmusait, és igyekszik azt a saját napirendje szerint manipulálni. Ne feledjük, hogy az amigdala nem képes önállóan felmérni a veszélyt - az emberi agyban a neokortexre támaszkodik, hogy meghozza ezeket az ítéleteket. Az a személy barát vagy ellenség, szerető vagy fenyegetés? Ezt csak a neokortex tudja eldönteni.

Amennyiben nem veszünk részt közvetlenül a halálos harcban és az élelemszerzésben, sikerült legalább részben szublimálnunk ősi késztetéseinket kreatívabb törekvések felé. Ezzel kapcsolatban a [következő fejezetben](#) a kreativitásról és a szerelemről lesz szó.

FEJEZET 6

TRANSZCENDENS KÉPESSÉGEK

Ez az én egyszerű vallásom. Nincs szükség diplomokra; nincs szükség bonyolult filozófiára. A saját agyunk, a saját szívünk a templomunk; a filozófia a kedvesség.

-A Dalai Láma

A kezem azért mozog, mert bizonyos erőket - elektromos, mágneses vagy bármilyen "idegerő" - az agyam hat rá. Ez az agyban tárolt idegerő valószínűleg - ha a Tudomány teljes lenne - a vér által az agyba juttatott kémiai erőkre lenne visszavezethető, és végső soron az általam elfogyasztott táplálékból és a belélegzett levegőből származna.

-Lewis Carroll

Érzelmi gondolataink szintén a neokortexben zajlanak, de az agy olyan ősi agyi régióktól kezdve, mint az amygdala, egészen néhány evolúciósan új agyi struktúráig, mint az orsó neuronok, amelyek a jelek szerint kulcsszerepet játszanak a magasabb szintű érzelmekben. Az agykéregben található szabályos és logikusan rekurzív struktúrákkal ellentétben az orsóneuronok rendkívül szabálytalan alakúak és összeköttetésekkel rendelkeznek. Ezek a legnagyobb neuronok az emberi agyban, amelyek annak teljes szélességén átívelnek. Mélyen összekapcsolódnak, több százezer kapcsolat köti össze a neokortex különböző részeit.

Mint korábban említettük, az insula segít feldolgozni az érzékszervi jeleket, de kulcsszerepet játszik a magasabb szintű érzelmekben is. Ebből a régióból származnak az orsósejtek. A funkcionális mágneses rezonancia képalkotó (fMRI) vizsgálatok kimutatták, hogy ezek a sejtek különösen aktívak, amikor az ember olyan érzelmekkel foglalkozik, mint a szerelem, a harag, a szomorúság és a szexuális vágy. Olyan helyzetek, amelyek erősen aktiválják őket, közé tartozik, amikor az alany ránéz a partnerére, vagy hallja, hogy a gyermeke sír.

Az orsósejtek hosszú idegszálakkal, úgynevezett apikális dendritekkel rendelkeznek, amelyek képesek távoli neokortikális régiókkal összeköttetést teremteni. Az ilyen "mély" összeköttetés, amelyben bizonyos neuronok számos régiót összekötnek, az evolúciós ranglétrán felfelé haladva egyre gyakrabban fordul elő. Nem meglepő, hogy az orsósejtek, amelyek részt vesznek az érzelmek és az erkölcsi ítéltételek kezelésében, rendelkeznek a kapcsolódás ezen formájával, tekintve, hogy a magasabb szintű érzelmi reakciók képesek különféle témákat és gondolatokat érinteni. Az agy számos más részével való kapcsolatuk miatt az orsósejtek által feldolgozott magas szintű érzelmekre valamennyi érzékelési és kognitív területünk hatással van. Fontos kiemelni, hogy ezek a sejtek nem végeznek racionális problémamegoldást, ezért nincs racionális kontrollunk a zenére adott reakcióink vagy a szerelembe esés felett. Az agy többi része azonban erősen részt vesz abban, hogy megpróbálja értelmezni rejtélyes, magas szintű érzelmeinket.

Az orsösejtek száma viszonylag kevés: mindössze körülbelül 80 000, ebből körülbelül 45 000 a jobb féltekében és 35 000 a bal féltekében. Ez az aránytalanság legalábbis az egyik oka annak a felfogásnak, hogy az érzelmi intelligencia a jobb agyfélteke sajátja, bár az aránytalanság szerény. A gorilláknak körülbelül, a bonobóknak körülbelül, a 2,100, csimpánzoknak pedig körülbelül Más1,800. emlősökből teljesen hiányoznak 16,000ezek a sejtek.

Az antropológusok úgy vélik, hogy az orsösejtek 10-15 millió évvel ezelőtt jelentek meg először a majmok és a hominidák (az ember elődei) még fel nem fedezett közös őseiben, és számuk 100 000 évvel ezelőtt gyorsan növekedett. Érdekes módon az orsösejtek nem léteznek az újszülött emberekben, hanem csak négy hónapos kor körül kezdenek megjelenni, és számuk egy és három éves kor között jelentősen megnő. Ugyanebben az időszakban fejlődik ki a gyermekek képessége, hogy erkölcsi kérdésekkel foglalkozzanak, és olyan magasabb szintű érzelmeket érzékeljenek, mint a szeretet.

Alkalmasság

Wolfgang Amadeus Mozart (1756-1791) ötéves korában írt egy menüettet. Hatévesen már fellépett Mária Terézia császárnőnek a bécsi császári udvarban. Harmincöt éves korában bekövetkezett haláláig hatszáz művet, köztük negyvenegy szimfóniát komponált, és széles körben az európai klasszikus hagyomány legnagyobb zeneszerzőjeként tartják számon. Mondhatnánk, hogy zenei adottságai voltak.

Mit jelent ez az elme mintafelismerési elméletének kontextusában? Nyilvánvaló, hogy az általunk adottságnak tekintett képességek egy része a nevelés, azaz a környezet és más emberek befolyása. Mozart zenészcsaládba született. Apja, Leopold, zeneszerző és a salzburgi érsek udvari zenekarának kapellmestere (szó szerint zenei vezetője) volt. Az ifjú Mozart elmerült a zenében, apja már hároméves korában elkezdte tanítani őt hegedülni és klaviatúrára (billentyűs hangszer). A környezeti hatások önmagukban azonban nem magyarázzák meg teljes mértékben Mozart zsenialitását. Egyértelműen van egy természeti összetevő is. Milyen formában jelenik meg ez? Ahogy a [4. fejezetben](#) írtam, a neokortex különböző régiói (a biológiai evolúció révén) bizonyos típusú mintákra optimalizálódtak. Bár a modulok alapvető mintafelismerő algoritmusai egységes az egész neokortexben, mivel bizonyos típusú minták hajlamosak bizonyos régiókon keresztül áramlani (például az arcok a fusiform gyruson keresztül), ezek a régiók jobban képesek lesznek a kapcsolódó minták feldolgozására. Számos paraméter határozza meg azonban, hogy az algoritmus ténylegesen hogyan valósul meg az egyes modulokban. Például milyen szoros egyezés szükséges ahhoz, hogy egy mintát felismerjünk? Hogyan módosul ez a küszöbérték, ha egy magasabb szintű modul jelzi, hogy az ő mintája "várható"? Hogyan veszik figyelembe a méretparamétereket?

Ezek és

egyéb tényezők a különböző régiókban eltérő módon lettek meghatározva, hogy előnyösek legyenek az egyes mintatípusok számára. A mesterséges intelligencia hasonló módszereivel végzett munkánk során ugyanezt a jelenséget vettük észre, és evolúciós szimulációkat használtunk e paraméterek optimalizálására.

Ha egyes régiók különböző típusú mintákra optimalizálhatók, akkor ebből az következik, hogy az egyes agyak is különböznek a tanulási, felismerési és létrehozási képességükben. Például egy agynak lehet veleszületett képessége a zenéhez, mivel jobban felismeri a ritmikus mintákat, vagy jobban megérti a harmóniák geometriai elrendezését. A tökéletes hangmagasság jelensége (a hangmagasság felismerésének és reprodukálásának képessége külső referencia nélkül), amely összefügg a zenei tehetséggel, úgy tűnik, genetikai alapokon nyugszik, bár a képességet fejleszteni kell, így valószínűleg a természet és a nevelés kombinációja. A tökéletes hangmagasság genetikai alapja valószínűleg a neokortexen kívül, a hallási információk előfeldolgozásában található, míg a tanult aspektus a neokortexben.

Vannak más készségek is, amelyek hozzájárulnak a kompetencia fokozatokhoz, legyenek azok rutinszerűek vagy legendásan zseniálisak. A neokortikális képességek - például a neokortex képessége, hogy elsajátítsa az amigdala által (rosszallás esetén) generált félelemjelzéseket - jelentős szerepet játszanak, csakúgy, mint az olyan tulajdonságok, mint az önbizalom, a szervezőkészség és a mások befolyásolásának képessége. Egy nagyon fontos képesség, amelyet korábban már említettem, a bátorság, hogy az ortodoxiával ellentétes elképzeléseket kövessünk. Az általunk zseniként számon tartott emberek mindig olyan módon folytatták saját mentális kísérleteiket, amelyet a társaik kezdetben nem értettek vagy nem értékelték. Bár Mozart még életében elismerést szerzett, a legtöbb elismerést később kapta. Koldusként halt meg, közönséges sírba temették, és csak két másik zenész jelent meg a temetésén.

Kreativitás

A kreativitás olyan drog, amely nélkül nem tudok élni.

-Cecil B. DeMille

A probléma sosem az, hogy hogyan juttassunk új, innovatív gondolatokat az elménkbe, hanem az, hogy hogyan juttassuk ki a régieket. Minden elme egy archaikus bútorokkal teli épület. Takarítsd ki az elméd egy sarkát, és a kreativitás azonnal meg fogja tölteni azt.

-Dee Hock

Az emberiség elég rideg tud lenni azokkal szemben, akiknek a szeme másképp látja a világot.

-Eric A. Burns

A kreativitás szinte minden problémát megoldhat. A kreatív aktus, a megszokás legyőzése az eredetiséggel, mindent legyőz.

-George Lois

A kreativitás egyik legfontosabb aspektusa a nagyszerű metaforák megtalálásának folyamata. -szimbólumok, amelyek valami mást képviselnek. A neokortex egy nagyszerű metaforagépezet, ami megmagyarázza, hogy miért vagyunk egyedülállóan kreatív faj. A neokortexünkben található körülbelül egymillió300 mintafelismerő közül mindegyik felismer és definiál egy mintát, és nevet ad neki, ami a neokortex mintafelismerő modulok esetében egyszerűen a mintafelismerőből kilépő axon, amely tüzelni fog, amikor az adott mintát megtaláljuk. Ez a szimbólum viszont aztán egy másik minta részévé válik. E minták mindegyike lényegében egy-egy metafora. A felismerők másodpercenként akár 100-szor is tüzelhetnek, így másodpercenként akár 30 milliárd metafora felismerésére is képesek vagyunk. Természetesen nem minden modul tüzel minden ciklusban - de azt mondhatjuk, hogy másodpercenként valóban több millió metaforát ismerünk fel.

Természetesen egyes metaforák jelentősebbek, mint mások. Darwin úgy vélte, hogy Charles Lyell felismerése arról, hogy a csordogáló víz nagyon fokozatos változásai hogyan tudnak nagy kanyonokat kialakítani, erőteljes metafora arra, hogy a több ezer generáción át tartó apró evolúciós változások csordogálása hogyan tud nagy változásokat létrehozni a fajok differenciálódásában. Az olyan gondolat kísérletek, mint amelyet Einstein használt a Michelson-Morley-kísérlet valódi értelmének megvilágítására, mind metaforák, abban az értelemben, hogy "valami másnak a reprezentánsának vagy szimbolikusának tekintett dolog", hogy egy szótári definíciót idézzek.

Látsz metaforákat 73Shakespeare Szonettjében?

Az évnék azt az időszakát láthatod bennem, Mikor sárga levelek, vagy egy sem, vagy kevés, lógnak...

Azokon az ágakon, melyek a hideg ellen remegnek, Csupán romos kórusok, ahol régen énekeltek az édes madarak. Bennem látod az ilyen nap alkonyát.

Mint napnyugta után a nyugaton elhalványul, Melyet a fekete éj idővel elvisz, A halál második énje, mely mindent nyugalomba zár. Bennem oly tüzet látsz izzani, mely ifjúságának hamvain hever,

Mint a halálos ágy, melyen el kell halnia, elfogyasztva azt, amiből táplálkozott.

Ez az, amit észreveszel, ami még erősebbé teszi a szerelmedet, Hogy szeresd azt a kútfit, amit hamarosan el kell hagynod.

Ebben a szonettben a költő kiterjedt metaforákat használ, hogy leírja a kor előrehaladását. Az ő kora olyan, mint a késő ős, "amikor sárga levelek, vagy egy sem, vagy kevés, lógnak". Hideg az idő, és a madarak már nem tudnak ülni az ágakon, amelyeket "csupaszmormókórusoknak" nevez. Az ő kora olyan, mint a szürkület, amikor "a napnyugta elhalványul nyugaton, amit a fekete éjszaka idővel elvisz". Ő a tűz maradványa, "mely ifjúságának hamvain fekszik". Valójában minden nyelv végső soron metafora, bár egyes kifejezései emlékezetesebbek, mint mások.

A metafora megtalálása egy minta felismerésének folyamata a részletekben és a kontextusban mutatkozó különbségek ellenére - ez egy olyan tevékenység, amelyet triviálisan végzünk életünk minden pillanatában. Az általunk jelentősnek tartott metaforikus ugrások általában a különböző tudományágak közti átjárásokban történnek. A kreativitásnak ez ellen a lényegi ereje ellen dolgozik azonban a tudományokban (és szinte minden más területen is) az egyre nagyobb specializálódás irányába mutató, mindent átható tendencia. Ahogy Norbert Wiener (1894-1964) amerikai matematikus írta a *kibernetika* című, korszakalkotó könyvében, amely az én születésem évében (1948) jelent meg:

A tudományos munkának vannak olyan területei, amint azt e könyv fő részében látni fogjuk, amelyeket a tiszta matematika, a statisztika, az elektrotechnika és a neurofiziológia különböző oldaláról vizsgáltak; amelyekben minden egyes fogalom külön nevet kap az egyes csoportoktól, és amelyekben fontos munkák háromszorosára vagy négyszeresére nőttek, míg más fontos munkák azért késnek, mert az egyik területen nem állnak rendelkezésre olyan eredmények, amelyek a következő területen már klasszikusnak számíthatnak. Ez ezek a határterületek kínálnak a leggazdagabb lehetőségeket a képzett kutatók számára. Ugyanakkor ezek a területek a legmegtagadhatóbbak a tömeges támadás és a munkamegosztás elfogadott technikáival szemben.

A saját munkám során a növekvő specializáció ellen az egyik módszer, amelyet a saját munkám során alkalmaztam, az, hogy összegyűjtöm a projekthez összegyűjtött szakértőket (például a beszédfelismerési munkámban beszédkutatók, nyelvészek, pszichoakusztikusok és mintafelismerési szakértők vettek részt, nem is beszélve a számítógépes tudósokról), és arra bátorítom őket, hogy tanítsák meg a csoportnak a saját technikáikat és terminológiájukat. Aztán kidobjuk az összes terminológiát, és kitaláljuk a sajátunkat. Változatlanul találunk olyan metaforákat az egyik területről, amelyek megoldják egy másik terület problémáit.

Az egér, amelyik megtalálja a menekülési útvonalat, amikor szembekerül a házimacskával - és ezt akkor is képes megtenni, ha a helyzet némileg eltér attól, amivel korábban valaha is találkozott -, kreatív. A mi kreativitásunk nagyságrendekkel nagyobb, mint az egéré - és sokkal több absztrakciós szintet foglal magában -, mivel sokkal nagyobb neokortexünk van, amely nagyobb hierarchiaszintekre képes. A nagyobb kreativitás elérésének egyik módja tehát az, hogy több neokortexet állítunk össze.

A rendelkezésre álló neokortex bővítésének egyik módja több ember együttműködése. Ez rutinszerűen egy problémamegoldó közösségben összegyűlt emberek

közötti kommunikációban keresztül valósul meg. A közelmúltban történtek erőfeszítések az online együttműködési eszközök használatára, hogy kihasználják a valós idejű együttműködés erejét,

amelyek a matematikában és más területeken is sikeresek voltak.¹

A következő lépés természetesen az lesz, hogy magát a neokortexet kibővítsük a nem biológiai megfelelőjével. Ez lesz a kreativitásunk végső aktuusa: a kreativitás képességének megteremtése. Egy nem biológiai neokortex végül is gyorsabb lesz, és gyorsan kereshetné azokat a fajta metaforákat, amelyek Darwint és Einsteint inspirálták. Szisztematikusan felfedezhetné az exponenciálisan bővülő tudásunk határai közötti összes átfedő határt.

Egyesek aggodalmukat fejezik ki, hogy mi lesz azokkal, akik nem kívánnak részt venni ebben az elmeterjesztésben. Rámutatnék, hogy ez a további intelligencia lényegében a felhőben (a számítógépek exponenciálisan bővülő hálózatában, amelyhez online kommunikáción keresztül csatlakozunk) lesz, ahol most a gépi intelligenciánk nagy része tárolódik. Amikor keresőmotort használunk, felismerjük a telefonunk beszédét, konzultálunk egy virtuális asszisztenssel, például Sirivel, vagy a telefonunkkal lefordítunk egy jelet egy másik nyelvre, az intelligencia nem magában az eszközben, hanem a felhőben van. A kibővített neokortexünk is ott kap helyet. Az, hogy ezt a kibővített intelligenciát közvetlen neurális kapcsolaton keresztül érzük el, vagy úgy, ahogy most tesszük - az eszközeinken keresztül interakcióba lépve vele -, önkényes különbségtétel. Véleményem szerint mindannyian kreatívabbak leszünk e mindent átható fejlesztés révén, függetlenül attól, hogy úgy döntünk-e, hogy közvetlen kapcsolatot választunk az emberiség kibővített intelligenciájával, vagy sem. Személyes, társadalmi, történelmi és kulturális emlékezetünk nagy részét már kiszerveztük a felhőbe, és végül ugyanezt fogjuk tenni a hierarchikus gondolkodásunkkal is.

Einstein áttörését nemcsak az eredményezte, hogy elmekísérleteken keresztül metaforákat alkalmazott, hanem az is, hogy bátran hitt e metaforák erejében. Hajlandó volt lemondani a hagyományos magyarázatokról, amelyek nem elégítették ki a kísérleteit, és hajlandó volt ellenállni a kortársai gúnyolódásának a metaforái által sugallt bizarr magyarázatokkal szemben. Ezeket a tulajdonságokat - a metaforákba vetett hitet és a meggyőződés bátorságát - nekünk is be kellene tudnunk programozni a nem biológiai neokortexünkbe.

Szerelem

Az elme tisztasága a szenvedély tisztaságát is jelenti; ezért a nagy és tiszta elme szenvedélyesen szeret és tisztán látja, amit szeret.

-Blaise Pascal

A szerelemben mindig van valami örültség. De az örületben mindig van némi értelem is.

-Friedrich Nietzsche

Ha már annyi mindent láttál az életből, mint én, nem fogod alábecsülni a megszállott szeretet erejét.

-Albus Dumbledore, J. K. Rowling, Harry Potter és a félvér herceg című könyvében.

Mindig is szerettem egy jó matematikai megoldást minden szerelmi problémára.

-Michael Patrick King, a Szex és New York "Vigyél ki a labdajátékra" című epizódjából.

Ha személyesen még nem is tapasztaltad meg az eksztatikus szerelmet, kétségtelenül hallottál már róla. Joggal mondhatjuk, hogy a világ művészetének jelentős részét, ha nem a többségét - történetek, regények, zene, tánc, festmények, televíziós műsorok és filmek - a szerelem történetei ihlették a legkorábbi szakaszában.

Nemrég a tudomány is beszállt a dologba, és már képesek vagyunk azonosítani azokat a biokémiai változásokat, amelyek akkor következnek be, amikor valaki szerelmes lesz. Dopamin szabadul fel, ami boldogság- és örömeztetést kelt. A noradrenalin szintje megugrik, ami szívdobogáshoz és általános feldobottság érzéséhez vezet. Ezek a vegyi anyagok a feniletilaminnal együtt elragadtatást, magas energiaszintet, összpontosított figyelmet, étvágytalanságot és általános vágyakozást váltanak ki a vágy tárgya után. Érdekes módon a londoni University College legújabb kutatása azt is kimutatta, hogy a szerotoninszint lecsökken, hasonlóan ahhoz, ami a kényszerbetegségben történik, ami összhangban van a következőkkel

a korai szerelem megszállottsága.² A dopamin és a noradrenalin magas szintje felelős a korai szerelem fokozott rövid távú figyelméért, eufóriájáért és sóvárgásáért.

Ha ezek a biokémiai jelenségek hasonlóan hangzanak, mint a harcolj vagy menekülj szindróma jelenségei, akkor azok is, kivéve, hogy itt valami vagy valaki felé futunk; sőt, egy cinikus azt mondaná, hogy inkább a veszély felé, mint a veszély elől. A változások teljes mértékben megfelelnek a függőségi viselkedés korai fázisaiban bekövetkező változásoknak is. A Roxy Music "Love Is the Drug" című dala elég pontosan leírja ezt az állapotot (bár a dal alánya a következő szerelmi adagját keresi). Az eksztatikus vallási élményekkel kapcsolatos tanulmányok is ugyanezeket a fizikai jelenségeket mutatják; azt lehet mondani, hogy az ilyen élményben részesülő személy beleszeret Istenbe vagy bármilyen spirituális kapcsolatba, amelyre összpontosít.

A korai romantikus szerelem esetében az ösztrogén és a tesztoszteron minden bizonnyal szerepet játszik a nemi vágy kialakulásában, de ha a szerelem egyetlen evolúciós célja a szexuális reprodukció lenne, akkor a folyamat romantikus aspektusára nem lenne szükség. Ahogy John William Money (1921-2006) pszichológus írta: "A vágy kéjes, a szerelem lírai".

A szerelem eksztatikus szakasza vezet a kötődési szakaszhoz és végül a hosszú távú kötődéshez. Vannak olyan vegyi anyagok, amelyek ezt a folyamatot is elősegítik, többek között az oxitocin és a vazopresszin. Nézzünk meg két rokon poloskafajt: a prérifarkas és a hegyi poloska.

Nagyjából azonosak, kivéve, hogy a préri egerészölyv rendelkezik oxitocin- és vazopresszin-receptorokkal, míg a hegyi egerészölyv nem. A prérifarkasok élethosszig tartó monogám kapcsolatokról ismertek, míg a hegyvidéki farkasok szinte kizárólag egyéjszakás kalandokat folytatnak. A poloskák esetében az oxitocin és a vazopresszin receptorok nagyjából meghatározóak a szerelmi életük jellegét illetően.

Bár ezek a vegyi anyagok az emberekre is hatással vannak, a neokortexünk átvette a vezető szerepet, mint minden másban, amit teszünk. A vakondoknak is van neokortexük, de ez bélyeg méretű és lapos, és éppen elég nagy ahhoz, hogy társat találjanak maguknak egy életre (vagy a hegyvidéki vakondok esetében legalább éjszakára), és más alapvető vakond viselkedési formákat hajtsanak végre. Nekünk, embereknek elegendő neokortexünk van ahhoz, hogy a Money által említett "lírai" kifejezésekkel foglalkozzunk.

Evolúciós szempontból maga a szeretet azért létezik, hogy kielégítse a neokortex szükségleteit. Ha nem lenne neokortexünk, akkor a vágy teljesen elegendő lenne a szaporodás biztosításához. A szerelem extatikus gerjesztése kötődéshez és érett szerelemhez vezet, és tartós kötődést eredményez. Ez viszont arra szolgál, hogy legalább a stabil környezet lehetőségét biztosítsa a gyermekek számára, amíg a saját neokortikájuk átveszi a felelősségteljes és képes felnőtté váláshoz szükséges kritikus tanuláson. A gazdag környezetben való tanulás eredendően része a neokortex módszerének. Valójában ugyanazok az oxitocin és vazopresszin hormonmechanizmusok játszanak kulcsszerepet a szülő (különösen az anya) és a gyermek közötti kritikus kötődés kialakításában.

A szerelem történetének túlsó végén a szeretett személy a neokortexünk fontos részévé válik. Az együtt töltött évtizedek után egy virtuális másik létezik a neokortexben, olyannyira, hogy minden lépését előre tudjuk, hogy mit fog mondani és tenni a szerelmünk. A neokortikális mintáinkat azokkal a gondolatokkal és mintákkal töltjük fel, amelyek azt tükrözik, hogy ki ő. Amikor elveszítjük ezt a személyt, szó szerint elveszítjük önmagunk egy részét. Ez nem *csak* egy metafora - a hatalmas mintafelismerők, amelyek tele vannak a szeretett személyt tükröző mintákkal, hirtelen megváltoztatják a természetüket. Bár értékes módszernek tekinthetjük őket arra, hogy életben tartsuk magunkban azt a személyt, az elvesztett szeretett személy hatalmas neokortikális mintázatait hirtelen az öröm kiváltóiból a gyász kiváltóivá válnak.

A szerelem evolúciós alapja és fázisai nem a teljes történet a mai világban. A szexet már nagyrészt sikerült megszabadítanunk biológiai funkciójától, hiszen lehet gyerekünk szex nélkül is, és biztosan lehet szexünk gyerek nélkül is. A szex túlnyomó többsége az érzéki és kapcsolati céljai miatt történik. És rutinszerűen szerelembe esünk a gyermeknevelésen kívüli más célokból is.

Hasonlóképpen, az ókorig visszanyúlóan a szerelmet és annak számtalan formáját ünneplő művészi kifejezésmódok hatalmas kiterjedése is öncélú. A transzcendens tudás e maradandó formáinak létrehozására való képességünk - legyen szó szerelemről vagy bármilyen másról - éppen az teszi egyedivé fajunkat.

A neokortex a biológia legnagyobb alkotása. A szerelemről szóló versek - és minden más alkotásunk - viszont a neokortexünk legnagyobb találmányai.

A BIOLÓGIAI IHLETÉSŰ DIGITÁLIS NEOKORTEX

Soha ne bízz semmiben, ami képes önállóan gondolkodni, ha nem látod, hol tartja az agyát.

-Arthur Weasley, J. K. Rowling, *Harry Potter és az azkabani fogoly című* könyvében.

Nem, nem érdekel, hogy erős agyat fejlesszek. Csak egy közepes agyat akarok, olyat, mint az Amerikai Telefon és Távíró Társaság elnöke.

-Alan Turing

Egy számítógép akkor érdemelné meg, hogy intelligensnek nevezzük, ha képes lenne megtéveszteni egy embert, hogy azt higgye, hogy ember.

-Alan Turing

Úgy vélem, hogy az évszázad végére a szóhasználat és az általános művelt vélemény annyira megváltozik, hogy az ember képes lesz úgy beszélni a gépi gondolkodásról, hogy nem számíthat ellentmondásra. - Alan Turing

Az anyapatkány akkor is fészket épít a kicsinyeinek, ha még soha életében nem látott másik patkányt.¹ Hasonlóképpen, egy pók hálót fon, egy hernyó megalkotja a saját gubóját, és egy hód gátat épít, még akkor is, ha soha egyetlen kortársuk sem mutatta meg nekik, hogyan kell elvégezni ezeket az összetett feladatokat. Ez nem jelenti azt, hogy ezek nem tanult viselkedések. Csakhogy ezek az állatok nem egyetlen életük során tanulták meg őket - hanem több ezer életen keresztül. Az állati viselkedés evolúciója valóban egy tanulási folyamat, de ez a faj tanulása, nem pedig az egyedé, és ennek a tanulási folyamatnak a gyümölcsei a DNS-ben vannak kódolva.

A neokortex evolúciójának jelentőségének megértéséhez gondoljunk arra, hogy a tanulás (hierarchikus tudás) folyamatát évezredekről hónapokra (vagy kevesebb hónapra) gyorsította fel. Még ha egy adott emlősfajban több millió állat nem is tudott megoldani egy (hierarchikus lépéseket igénylő) problémát, akkor is csak egy kellett ahhoz, hogy véletlenül rábukkanjon a megoldásra. Ezt az új módszert aztán lemásolták, és exponenciálisan terjedt a populációban.

Most abban a helyzetben vagyunk, hogy a biológiai intelligenciáról a nem biológiai intelligenciára való áttéréssel ismét ezerszer vagy milliószorosára gyorsíthatjuk fel a tanulási folyamatot. Ha egy digitális neokortex egyszer megtanul egy készséget, akkor percek vagy

akár másodpercek alatt át tudja adni ezt a tudást. Egy példa a sok közül: az első cégemnél, a Kurzweil Computer Productsnál (ma Nuance Speech Technologies), amelyet én alapítottam, éveket 1973-töltöttünk azzal, hogy egy sor kutató számítógépet betanítottunk arra, hogy szkennelt dokumentumokból felismerjék a nyomtatott betűket, ezt a technológiát omni-font (bármilyen betűtípusú) optikai karakterfelismerésnek (OCR) nevezik. Ezt a technológiát már majdnem negyven éve folyamatosan fejlesztik, a Nuance jelenlegi terméke az OmniPage. Ha azt szeretnénk, hogy a számítógépünk felismerje a nyomtatott betűket, nem kell éveket töltenünk a betűfelismerés betanításával, ahogyan mi tettük - egyszerűen letölthetjük a kutató számítógépek által már megtanult, kialakult mintákat szoftver formájában. Az 1980-
as években kezdtünk bele a beszédfelismerésbe, és ez a technológia, amely szintén több évtizede folyamatos fejlesztés alatt áll, része a Sirinek. Itt is másodpercek alatt letöltheti a kutatószámítógépek által hosszú évek alatt megtanult, továbbfejlesztett mintákat.

Végső soron egy olyan mesterséges neokortexet fogunk létrehozni, amely az emberi megfelelőjének teljes skálájával és rugalmasságával rendelkezik. Gondoljunk csak az előnyökre. Az elektronikus áramkörök több milliószor gyorsabbak, mint a biológiai áramköreink. Eleinte az egész sebességnövekedést arra kell majd fordítanunk, hogy kompenzáljuk a számítógépek relatív párhuzamossági hiányát, de végül a digitális neokortex sokkal gyorsabb lesz, mint a biológiai változat, és csak tovább fog nőni a sebessége.

Amikor a saját neokortexünket egy szintetikus változattal bővítjük, nem kell majd azon aggódnunk, hogy fizikailag mennyi neokortex fér el a testünkben és az agyunkban, mivel a neokortex nagy része a felhőben lesz, mint a ma használt számítástechnika nagy része. Korábban úgy becsültem, hogy a biológiai neokortexünkben körülbelül 300 millió mintafelismerő van. Ennyit még a nagy homlok evolúciós újításával is be tudnánk préselni a koponyánkba, és a neokortex a rendelkezésre álló hely körülbelül százalékát80 foglalná el. Amint elkezdünk a felhőben gondolkodni, nem lesznek természetes korlátok.

-mintázatfelismerők milliárdjait vagy trillióit leszünk képesek használni, alapvetően bármit, amire szükségünk van, és amit a gyorsuló megtérülés törvénye minden egyes időpontban biztosítani tud.

Ahhoz, hogy egy digitális neokortex megtanuljon egy új készséget, még mindig sok ismétléses oktatásra lesz szüksége, ahogyan egy biológiai neokortexnek is, de ha egyszer egy digitális neokortex valahol és valamikor megtanul valamit, akkor ezt a tudást késedelem nélkül meg tudja osztani minden más digitális neokortexszel. Mindannyiunknak lehetnek saját privát neokortex-kiterjesztőink a felhőben, ahogyan ma is megvannak a saját személyes adattárolóink.

Végül, de nem utolsósorban, képesek leszünk az intelligenciánk digitális részét is átátmasztani. Amint láttuk, nem csak metafora az a kijelentés, hogy a neokortexünkben információ található, és ijesztő belegondolni, hogy ma ezekből az információkból semmi sincs lementve. Van persze egy módja annak, hogy az agyunkban lévő információk egy részét lejegyezzük. Az a képesség, hogy gondolkodásunk legalább egy részét olyan adathordozóra tudjuk átültetni, amely túlélheti biológiai testünket, hatalmas előrelépés volt, de az agyunkban lévő adatok nagy része továbbra is sérülékeny marad.

Agy-szimulációk

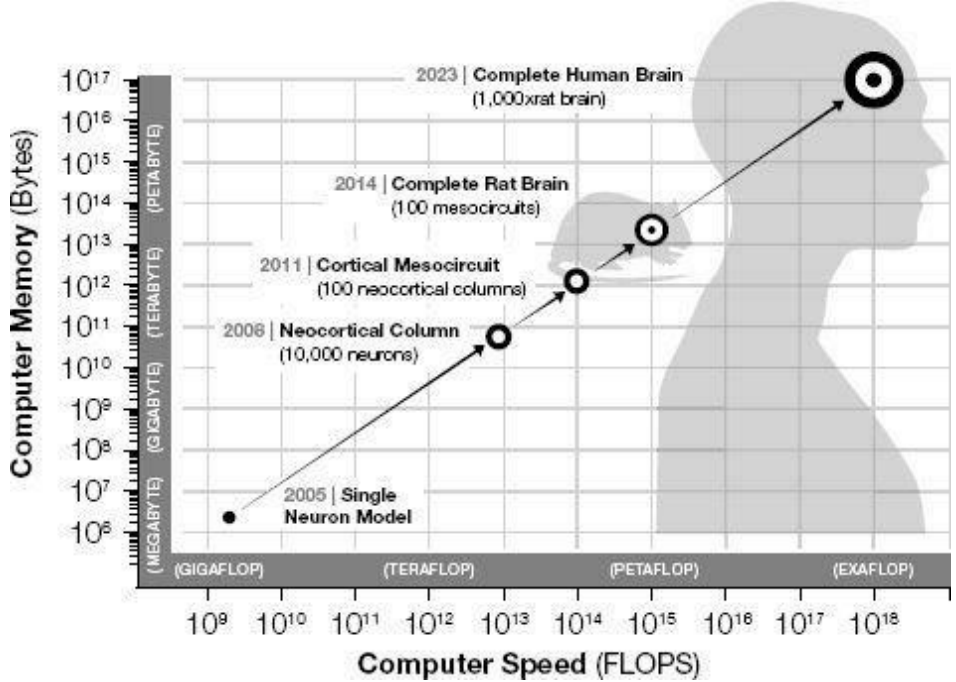
A digitális agy építésének egyik megközelítése a biológiai agy pontos szimulációja. A Harvard agytudományok doktori hallgatója, David Dalrymple (1991-es születésű) például egy fonálféreg (egy körömféreg) agyának szimulálását tervezi.² Dalrymple azért választotta a fonálférget, mert viszonylag egyszerű idegrendszere van, amely körülbelül 300 neuronból áll, és amelyet a molekulák nagyon részletes szintjén tervez szimulálni. Emellett számítógépes szimulációt készít a testéről és a környezetéről is, hogy a virtuális fonálféreg (virtuális) táplálékra vadászhasson, és elvégezhesse azokat a dolgokat, amelyekben a fonálféreg jó. Dalrymple szerint valószínűleg ez lesz az első teljes agyfeltöltés egy biológiai állatból egy virtuális világban élő virtuális állatba. Az ő szimulált fonálféregéhez hasonlóan az is vita tárgya, hogy még a biológiai fonálféreg is tudatosak-e. Bár az evésért, a táplálék megemésztéséért, a ragadozók elkerüléséért és a szaporodásért folytatott küzdelmük során vannak olyan tapasztalataik, amelyeknek tudatában lehetnek.

A spektrum másik végén Henry Markram Blue Brain Projectje az emberi agy szimulálását tervezi, beleértve a teljes neokortexet, valamint az olyan régi agyi régiókat, mint a hippokampusz, az amigdala és a kisagy. Tervezett szimulációi különböző részletességgel épülnek majd fel, egészen a molekuláris szintű teljes szimulációig. Amint arról a [4. fejezetben](#) beszámoltam, Markram felfedezett egy több tucat neuronból álló kulcsmodult, amely újra és újra megismétlődik a neokortexben, ami azt bizonyítja, hogy a tanulás ezekben a modulokban, és nem az egyes neuronokban történik.

Markram fejlődése exponenciális ütemben növekszik. Egyetlen neuront szimulált 2005-ben, a projekt elindításának évében. Csapata 2008-ban egy patkányagy teljes, 10 000 neuronból álló neokortikális oszlopát szimulálta. Ez 2011-re 100 oszlopra, összesen egymillió sejtre bővült, amit ő mezőkörnek nevez. Markram munkájával kapcsolatban az egyik ellentmondás az, hogy hogyan lehet ellenőrizni, hogy a szimulációk pontosak-e. Ehhez ezeknek a szimulációknak tanulást kell demonstrálniuk, amelyet alább tárgyalok.

Tervei szerint 2014-re egy teljes patkányagyat szimulál, amely 100 mezocirkuláris áramkörből áll, összesen 100 millió neuronnal és körülbelül egy trillió szinapszissal. Az oxfordi TED-konferencián 2009 tartott előadásában Markram azt mondta: "Nem lehetetlen megépíteni egy emberi agyat, és 10 éven belül meg tudjuk csinálni".³ Legfrissebb célkitűzése a teljes agyszimulációra 2023.⁴

Markram és csapata a modelljüket valódi neuronok részletes anatómiai és elektrokémiai elemzéseire alapozzák. Egy általuk készített automatizált eszközzel, az úgynevezett patch-clamp robottal mérik az egyes neuronokon belüli elektrokémiai aktivitásért felelős specifikus ioncsatornákat, neurotranszmittereket és enzimeket. Automatizált rendszerük Markram szerint hat hónap alatt harminc évnyi elemzést volt képes elvégezni. Ezekből az elemzésekből vették észre a "Lego memória" egységeket, amelyek a neokortex alapvető funkcionális egységei.

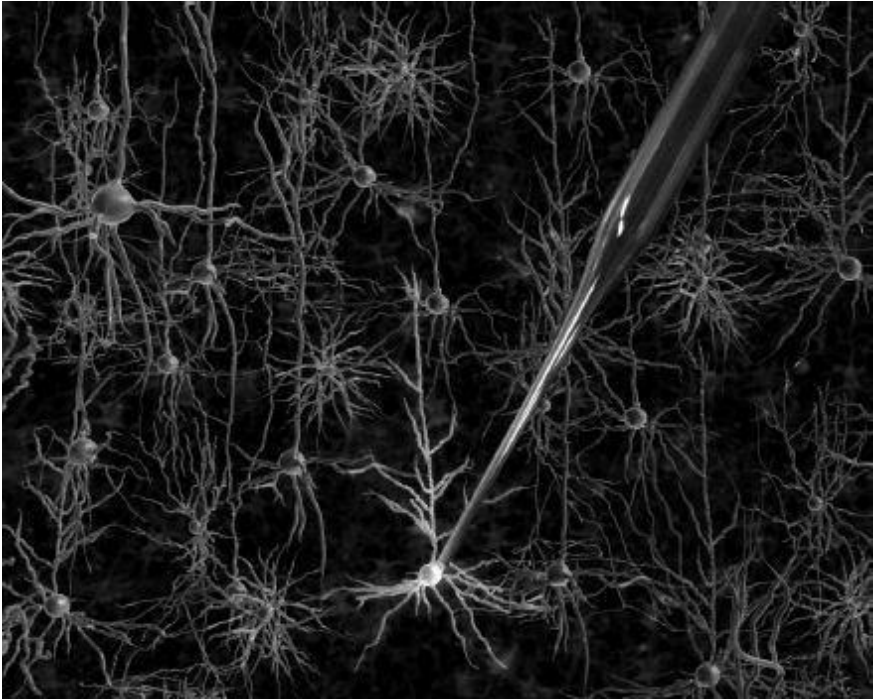


A Blue Brain agyszimulációs projekt tényleges és tervezett előrehaladása.

A robotizált tapasztorzás technológiájához jelentősen hozzájárult Ed Boyden, az MIT idegkutatója, Craig Forest, a Georgia Tech gépészmérnök professzora és Suhasa Kodandaramaiah, Forest végzős hallgatója. Bemutattak egy olyan egy mikrométeres pontosságú automatizált rendszert, amely képes az idegszövetek pásztázását nagyon közről elvégezni anélkül, hogy az idegsejtek érzékeny membránjait károsítaná. "Ez olyasvalami, amit egy robot meg tud csinálni, amit egy ember nem" - kommentálta Boyden.

Visszatérve Markram szimulációjára, miután szimulált egy neokortikális oszlopot, Markramot idézve azt mondta: "Most már csak fel kell méreteznünk".⁵ A méretezés minden bizonnyal az egyik nagy tényező, de van egy másik kulcsfontosságú akadály is, ami a tanulás. Ha a Blue Brain Project agya "beszélni fog, intelligenciával fog rendelkezni, és úgy fog viselkedni, mint a

egy ember" - így jellemezte Markram a célját egy 2009-es BBC-interjúban -, akkor a szimulált neokortexnek elegendő tartalommal kell rendelkeznie ahhoz, hogy ezeket a feladatokat elvégezze.⁶ Mint azt bárki tanúsíthatja, aki próbált már beszélgetést folytatni egy újszülöttel, rengeteget kell tanulni ahhoz, hogy ez megvalósítható legyen.



Az MIT-n és a Georgia Tech-en kifejlesztett patch-clamping robot hegye az idegszövetet pásztázza.

Ezt két nyilvánvaló módon lehet megvalósítani egy szimulált agyban, például a Blue Brainben. Az egyik az lenne, hogy az agy megtanulja ezt a tartalmat úgy, ahogyan az emberi agy teszi. Kezdhethé úgy, mint egy újszülött emberi csecsemő, akinek veleszületett képessége van a hierarchikus tudás elsajátítására, és bizonyos átalakítások előre beprogramozva vannak az érzékszervi előfeldolgozó régióiban. De annak a tanulásnak, ami egy biológiai csecsemő és egy beszélgetést folytatni képes emberi személy között zajlik, hasonló módon kellene megtörténnie a nem biológiai tanulásban. Ezzel a megközelítéssel az a probléma, hogy egy olyan agy, amelyet olyan részletességgel szimulálnak, mint amilyen részletességgel a Blue Brain esetében várható, várhatóan legalább a 2020-as évek elejéig nem fog valós időben futni. Még a valós idejű futtatás is túl lassú lenne, hacsak a kutatók nem hajlandóak egy-két évtizedet várni arra, hogy elérjék a felnőtt emberrel való szellemi egyenrangúságot, bár a valós idejű teljesítmény folyamatosan gyorsulni fog, ahogy a számítógépek ára/teljesítménye folyamatosan nő.

A másik megközelítés az, hogy veszünk egy vagy több biológiai emberi agyat, amelyek már elegendő tudást szereztek ahhoz, hogy értelmes nyelven társalognanak, és egyébként érett módon viselkedjenek, és átmásoljuk neokortikális mintáikat a szimulált agyba. Ezzel a módszerrel az a probléma, hogy egy ilyen feladat gyors és teljes körű elvégzéséhez megfelelő térbeli és időbeli felbontású és sebességű, nem invazív és

roncsolásmentes szkennelési technológiára van szükség. Nem számítanék arra, hogy egy ilyen "feltöltési" technológia a 2040-es évekig elérhető lesz. (Az agy szimulációjának számítási igénye ekkorra már nem lenne elegendő.

pontossági fok, amely becslésem szerint másodpercenként 10^{19} számítás lesz, az előrejelzéseim szerint a 2020-as évek elejére elérhető lesz egy szuperszámítógépben; a szükséges roncsolásmentes agyszkenelési technológiák azonban hosszabb időt vesznek igénybe.).

Van egy harmadik megközelítés is, amelyről úgy vélem, hogy a Blue Brainhez hasonló szimulációs projekteknek kell követniük. A molekuláris modelleket egyszerűsíthetjük azért, hogy különböző szintű funkcionális megfelelőket hozunk létre, a saját funkcionális algoritmikus módszeremtől kezdve (ahogyan azt ebben a könyvben leírtam) a teljes molekuláris szimulációkhoz közelebb álló szimulációkig. A tanulás sebessége ezáltal százszorosára vagy ezerszeresére növelhető az alkalmazott egyszerűsítés mértékétől függően. A szimulált agy számára (a funkcionális modellt használva) olyan oktatási programot lehet kidolgozni, amelyet viszonylag gyorsan meg tud tanulni. Ezután a teljes molekuláris szimulációval helyettesíthető az egyszerűsített modell, miközben továbbra is felhasználható a felhalmozott tanulás. Ezután a teljes molekuláris modellel sokkal lassabban szimulálhatjuk a tanulást.

Dharmendra Modha amerikai informatikus és IBM-es kollégái létrehozták az emberi vizuális neokortex egy részének sejtről sejtre történő szimulációját, amely egymilliárd 1.6 virtuális neuronból és 9 trillió szinapsziszból áll, ami egy macska neokortexének felel meg. A valós időnél 100-szor lassabban fut a 147 456 processzorból álló IBM BlueGene/P szuperszámítógépen. A munka elnyerte az Association for Computing Machinery Gordon Bell-díját.

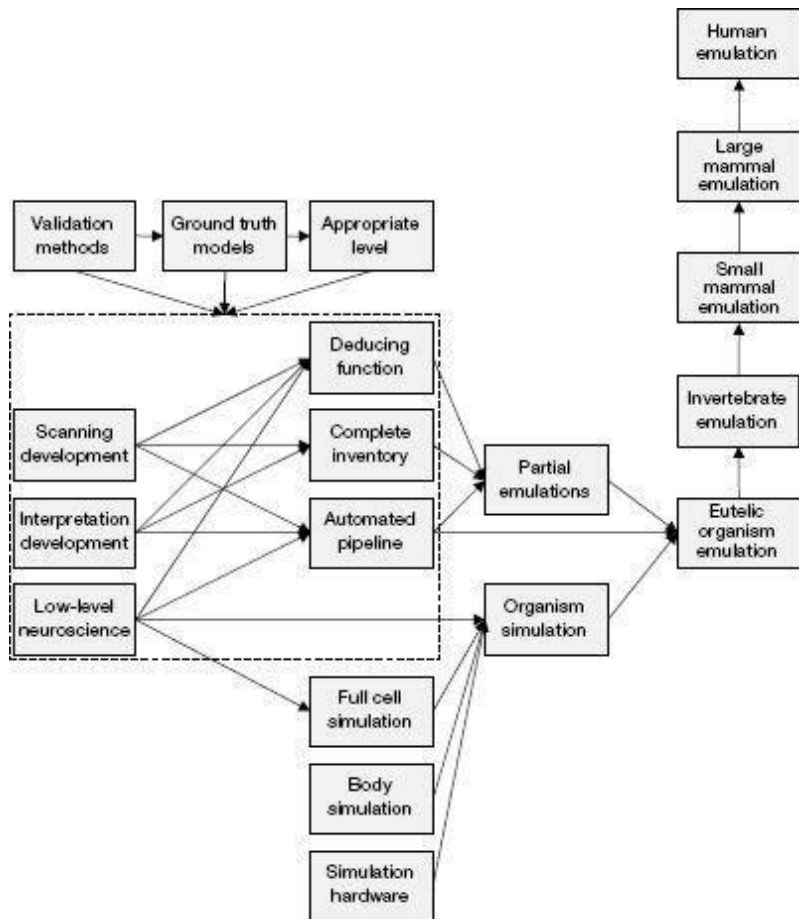
Az olyan agyszimulációs projektek, mint a Blue Brain és Modha neokortex-szimulációi, kifejezetten egy funkcionális modell finomítására és megerősítésére irányulnak. Az emberi szintű mesterséges intelligencia alapvetően az ebben a könyvben tárgyalt funkcionális algoritmikus modellt fogja használni. A molekuláris szimulációk azonban segíteni fognak abban, hogy tökéletesítsük ezt a modellt, és hogy teljes mértékben megértsük, mely részletek fontosak. Az 1980-as és 1990-es években a beszédfelismerő technológia fejlesztése során képesek voltunk finomítani az algoritmusainkat, miután megértettük a hallóideg és a hallókéreg korai részei által végzett tényleges átalakításokat. Még ha a funkcionális modellünk tökéletes is lenne, annak megértése, hogy pontosan hogyan valósul meg a biológiai agyunkban, fontos ismereteket fog feltárni az emberi működésről és működési zavarokról.

Részletes adatokra lesz szükségünk a tényleges agykról, hogy biológiailag megalapozott szimulációkat készíthessünk. Markram csapata saját adatokat gyűjt. Vannak nagyszabású projektek az ilyen típusú adatok összegyűjtésére és a tudósok számára általánosan hozzáférhetővé tételére. A New York-i Cold Spring Harbor Laboratory például 500 terabájtnyi adatot gyűjtött össze egy emlős (egér) agyának szkennelésével, amelyet júniusban tettek hozzáférhetővé. 2012. Projektjük lehetővé teszi, hogy a felhasználó hasonló módon fedezze fel az agyat, mint ahogy a Google Earth segítségével a bolygó felszínét. A

felhasználó mozoghat az egész agyban, és ráközelíthet az egyes neuronok és kapcsolataik megtekintéséhez. Kiemelhetünk egyetlen kapcsolatot, majd követhetjük annak útját az agyban.

A National Institutes of Health tizenhat szekciója összefogott, és 38,5 millió dollárral támogatta a Human Connectome Project nevű nagyszabású kezdeményezést.⁷ A St. Louis-i Washington Egyetem, a Minnesotai Egyetem vezetésével,

A Harvard Egyetem, a Massachusetts General Hospital és a Los Angeles-i Kaliforniai Egyetem közös projektje az emberi agyban lévő kapcsolatok hasonló háromdimenziós térképének létrehozására irányul. A projektben különböző nem invazív szkennelési technológiákat használnak, többek között az MRI új formáit, a magnetoencefalográfiát (az agy elektromos aktivitása által létrehozott mágneses mezők mérése) és a diffúziós traktográfiát (az agyi rostkötegek útvonalainak nyomon követésére szolgáló módszer). Amint arra a [10. fejezet](#)ben rámutatok, az agy nem invazív szkennelésének térbeli felbontása exponenciális ütemben javul. Van J. Wedeen és munkatársai Massachusetts General Hospitalban végzett kutatása, amely a neokortex vezetékezésének rendkívül szabályos, rácsszerű szerkezetét mutatja, és amelyet a [4. fejezetben](#) ismertettem, e projekt egyik korai eredménye.



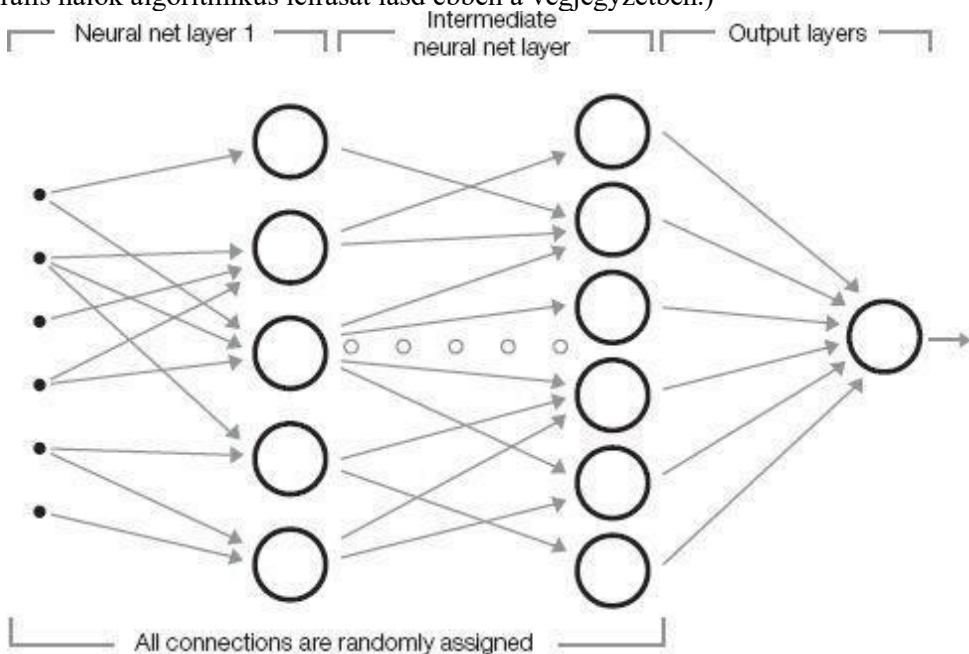
A teljes agyi emuláció vázlatja: Anders Sandberg és Nick Bostrom: Egy útiter.

Neurális hálók

1964-ben, tizenhat évesen írtam Frank Rosenblattnak (1928- 1971), a Cornell Egyetem professzorának, és érdeklődtem egy Mark 1 Perceptron nevű gépről. Négy évvel korábban alkotta meg, és úgy írta le, hogy agyszerű tulajdonságokkal rendelkezik. Meghívott, hogy látogassam meg, és próbáljam ki a gépet.

A Perceptront az idegsejtek elektronikus modelljeiből építette fel, amelyekről azt állította, hogy azok elektronikus modelljei. A bemenet két dimenzióban elrendezett értékekből állt. A beszéd esetében az egyik dimenzió a frekvenciát, a másik az időt jelentette, így minden egyes érték egy adott időpontban egy frekvencia intenzitását jelentette. A képek esetében minden egyes pont egy képpont volt egy kétdimenziós képen. Egy adott bemenet minden egyes pontja véletlenszerűen kapcsolódott a szimulált neuronok első rétegének bemeneteihez. Minden kapcsolathoz szinaptikus erősség tartozott, amely a fontosságát reprezentálta, és amelyet kezdetben véletlenszerű értékre állítottunk be. Minden neuron összeadta a hozzá érkező jeleket. Ha az összesített jel meghaladott egy bizonyos küszöbértéket, a neuron tüzelni kezdett, és jelet küldött a kimeneti kapcsolatára; ha az összesített bemeneti jel nem haladta meg a küszöbértéket, a neuron nem tüzel, és a kimenete nulla volt. Az egyes neuronok kimenete véletlenszerűen kapcsolódott a következő rétegben lévő neuronok bemenetéhez. A Mark 1 Perceptron három réteggel rendelkezett, amelyek különböző konfigurációkban szervezhetők voltak. Például az egyik réteg visszatáplálhatott egy korábbi rétegre. A legfelső rétegben egy vagy több, szintén véletlenszerűen kiválasztott neuron kimenete adta a választ.

(A neurális hálók algoritmikus leírását lásd ebben a végjegyzetben.)²



Mivel a neurális háló vezetékzése és a szinaptikus súlyok kezdetben véletlenszerűen vannak beállítva, a nem képzett neurális háló válaszai is véletlenszerűek. A neurális háló kulcsa tehát az, hogy meg kell tanulnia a tárgyát, akár csak az emlősök agyának, amelyről állítólag mintázták. A neurális háló kezdetben tudatlan; a tanítója - aki lehet egy ember, egy számítógépes program, vagy esetleg egy másik, érettebb neurális háló, amely már megtanulta a leckét - jutalmazza a tanuló neurális hálót, ha helyes kimenetet generál, és bünteti, ha nem. Ezt a visszajelzést viszont a tanuló neuronháló arra használja fel, hogy az egyes interneuronális kapcsolatok erősségét beállítsa. A helyes válasszal összhangban lévő kapcsolatok megerősödnek. Azok, amelyek a rossz választ támogatják, gyengülnek.

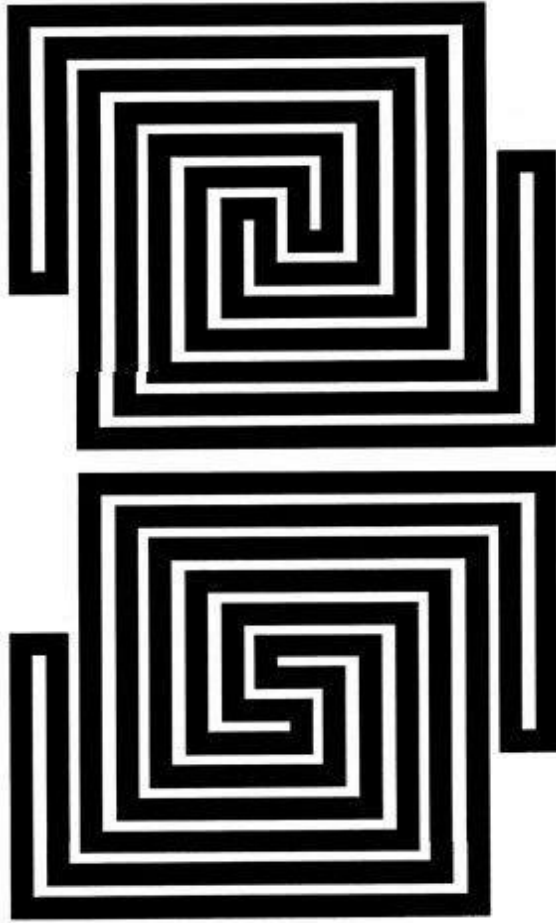
Idővel a neurális háló úgy szerveződik, hogy edzés nélkül is helyes válaszokat adjon. Kísérletek kimutatták, hogy a neurális hálók még megbízhatatlan tanárokkal is képesek megtanulni a tárgyukat. Ha a tanárnak csak az esetek 60 százalékában van igaza, a tanuló neurális háló akkor is 100 százalékot megközelítő pontossággal tanulja meg a leckét.

A Perceptron által megtanulható anyagok körének korlátai azonban hamar nyilvánvalóvá váltak. Amikor 1964-ben meglátogattam Rosenblatt professzort, egyszerű módosításokat próbáltam ki a bemeneten. A rendszert úgy állítottuk be, hogy felismerje a nyomtatott betűket, és elég pontosan fel is ismerte azokat. Meglehetősen jó munkát végzett az autoasszociációval (vagyis akkor is felismerte a betűket, ha egyes részeket eltakartam), de kevésbé jól teljesített az invariáns (vagyis a méret- és betűtípus-változásokra való általánosítással, ami összezavarta).

Az 1960-as évek utolsó felében ezek a neurális hálók hatalmas népszerűsége tettek szert, és a "konnekcionizmus" a mesterséges intelligencia területének legalább a felét átvette. A mesterséges intelligencia hagyományosabb megközelítése eközben közvetlen kísérleteket tartalmazott konkrét problémák megoldásának programozására, például arra, hogy hogyan ismerjük fel a nyomtatott betűk invariáns tulajdonságait.

Egy másik személy, akit 1964-ben meglátogattam, Marvin Minsky (született 1927-ben), a mesterséges intelligencia egyik alapítója volt. Annak ellenére, hogy az 1950-es években maga is úttörő munkát végzett a neurális hálókkal kapcsolatban, őt is érintette a technika iránti nagy érdeklődés hulláma. A neurális hálók vonzerejének része volt, hogy állítólag nem igényelnek programozást - a problémák megoldását maguktól megtanulják. 1965-ben kerültem be az MIT-re diákként, és Minsky professzor volt a mentorom, és osztottam a szkepticizmusát a "konnekcionizmus" iránti örületet illetően.

1969-ben Minsky és Seymour Papert (született 1928-ban), az MIT Mesterséges Intelligencia Laboratóriumának két társalapítója könyvet írt *Perceptronok* címmel, amelyben egyetlen központi tételt mutatott be: nevezetesen, hogy egy perceptron eleve képtelen meghatározni, hogy egy kép összefügg-e vagy sem. A könyv nagy vihart kavart. Annak megállapítása, hogy egy kép összefügg-e vagy sem, olyan feladat, amelyet az emberek nagyon könnyen el tudnak végezni, és az is egyszerű folyamat, hogy egy számítógépet úgy programozunk, hogy elvégezze ezt a megkülönböztetést. Azt, hogy a Perceptronok erre nem voltak képesek, sokan végzetes hibának tartották.



Két kép Marvin Minsky és Seymour Papert *Perceptrons* című könyvének borítójáról. A felső kép nem összefüggő (azaz a sötét terület két, nem összefüggő részből áll). Az alsó kép összefüggő. Ezt egy ember könnyen meg tudja állapítani, ahogy egy egyszerű szoftverprogram is. Egy előrecsatolt perceptron, mint például Frank Rosenblatt Mark Perceptronja, nem képes ezt a megállapítást megtenni.

A *Perceptrons*-t azonban széles körben úgy értelmezték, hogy többet jelent, mint amennyit valójában jelentett. Minsky és Papert tétele csak a neurális hálók egy bizonyos típusára, a feedforward neurális hálókra vonatkozott (ebbe a kategóriába tartozik a Rosenblatt-féle Perceptron is); a neurális hálók más típusaira nem vonatkozott ez a korlátozás. Ennek ellenére a könyvnek sikerült a neurális hálók kutatásának finanszírozását az 1970-es években nagyrészt megszüntetnie. A terület az 1980-as években tért vissza a biológiai neuronok realisabbnak vélt modelljeinek használatára tett kísérletekkel, amelyek

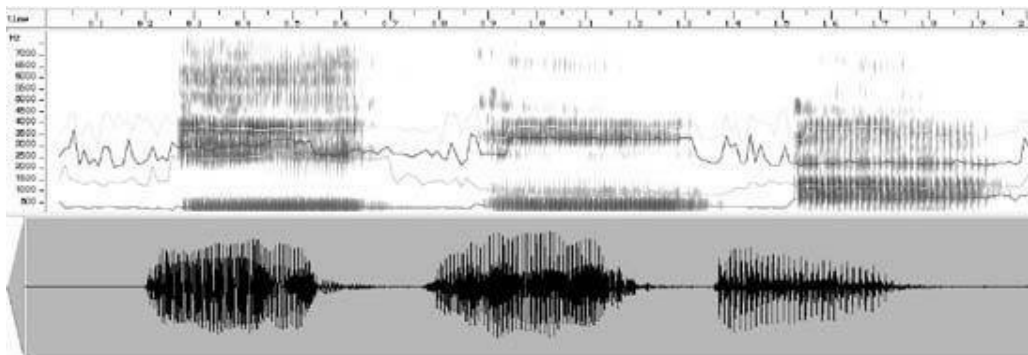
elkerülték a Minsky-Papert Perceptron-tételben foglalt korlátozásokat. Mindazonáltal a neokortex erősségének egyik kulcsa, az invariációs probléma megoldásának képessége olyan képesség volt, amely az újraéledő konneccionista terület számára is megfoghatatlan maradt.

Ritkított kódolás: Vektor kvantálás

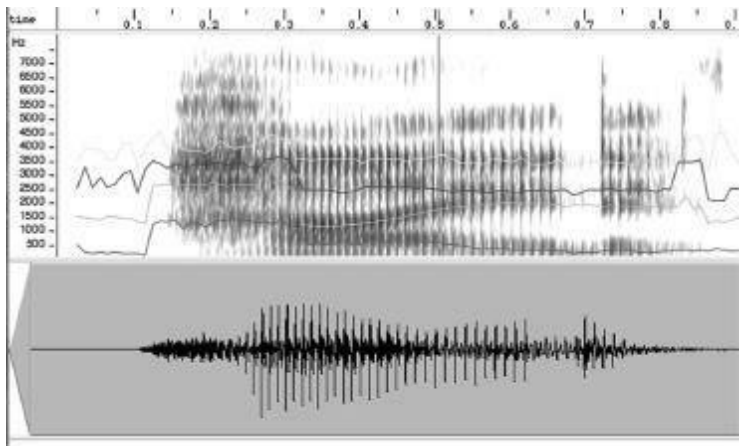
Az 1980-as évek elején egy másik klasszikus mintafelismerési problémával kezdtem el foglalkozni: az emberi beszéd megértésével. Először hagyományos mesterséges intelligencia-megközelítéseket alkalmaztunk, közvetlenül programozva a beszéd alapvető egységeire - a fonémákra - vonatkozó szakértői tudást, valamint a nyelvészekről származó szabályokat arról, hogy az emberek hogyan fűzik egymás mellé a fonémákat, hogy szavakat és kifejezéseket alkossanak. Minden egyes fonémának jellegzetes frekvenciamintázata van. Tudtuk például, hogy az olyan magánhangzókat, mint az "e" és az "ah", bizonyos rezonáns frekvenciák, úgynevezett formantsok jellemzik, és minden egyes fonémára jellemző arányuk van. Az olyan szibiláns hangokat, mint a "z" és az "s", több frekvencián átívelő hangkitörés jellemzi.

A beszédet hullámformaként rögzítettük, amelyet aztán egy sor frekvenciaszűrő segítségével több (hangmagasságként érzékelt) frekvenciasávra alakítottunk át. Ennek az átalakításnak az eredményét vizualizálni lehetett, és spektrogramnak neveztük el (lásd [136. oldal](#)).

A szűrőbank azt másolja, amit az emberi csiga csinál, ami a hang biológiai feldolgozásának első lépése. A szoftver először a fonémákat azonosította a frekvenciák megkülönböztető mintázatai alapján, majd a szavakat a fonémák jellegzetes sorozatainak azonosítása alapján.



Három magánhangzó spektrogramja. Balról jobbra: [i], mint az "appreciate", [u], mint az "acoustic", és [a], mint az "ah". Az Y tengely a hang frekvenciáját jelöli. Minél sötétebb a sáv, annál több akusztikus energia van az adott frekvencián.



Egy spektrogram egy személyről, aki azt mondja, hogy "elrejtőzni". A vízszintes vonalak a formánsokat mutatják, amelyek különösen nagy energiájú, tartós frekvenciák.¹⁰

Az eredmény részben sikeres volt. Be tudtuk képezni a készüléket egy adott személyre vonatkozó minták megtanulása egy közepes méretű, több ezer szóban mért szókincs segítségével. Amikor több tízezer szó felismerésére, több beszélő kezelésére és teljesen folyamatos beszédre (azaz a szavak közötti szünetek nélküli beszédre) tettünk kísérletet, beleütköztünk az invariancia problémájába. A különböző emberek ugyanazt a fonémát különbözőképpen ejtették ki - például az egyik ember "e" fonémája úgy hangzott, mint a másiké "ah". Még ugyanaz a személy is következtelenül beszélt egy adott fonémát. Egy fonéma mintázatát gyakran befolyásolták a közelben lévő más fonémák. Sok fonéma teljesen kimaradt. A szavak kiejtése (vagyis az, hogy a fonémák hogyan állnak össze szavakká) szintén nagyon változó volt, és függött a kontextustól. Az általunk programozott nyelvi szabályok kezdtek összeomlani, és nem tudtak lépést tartani a beszélt nyelv rendkívüli változékonyságával.

Akkoriban vált világossá számomra, hogy az emberi minta- és fogalomfelismerés lényege hierarchiakon alapul. Ez minden bizonnyal nyilvánvaló az emberi nyelv esetében, amely a struktúrák bonyolult hierarchiáját alkotja. De mi az az elem, amely a struktúrák alapját képezi? Ez volt az első kérdés, amelyet megvizsgáltam, amikor a teljesen normális emberi beszéd automatikus felismerésének módját kerestem.

A hang a levegő rezgéseként jut be a fülbe, és a csigaházban található mintegy 3000 belső szőrsejt több frekvenciasávra alakítja át. Minden egyes szőrsejt egy adott frekvenciára van hangolva (megjegyezzük, hogy a frekvenciákat hangként érzékeljük), és mindegyik frekvenciaszűrőként működik, jelet bocsát ki, amikor a rezonanciafrekvenciáján vagy annak közelében hang hallatszik. Az emberi csigahártyát elhagyva a hangot így körülbelül 3000 különálló jel képviseli, amelyek mindegyike egy-egy keskeny frekvenciasáv időben változó intenzitását jelzi (e sávok között jelentős átfedés van).

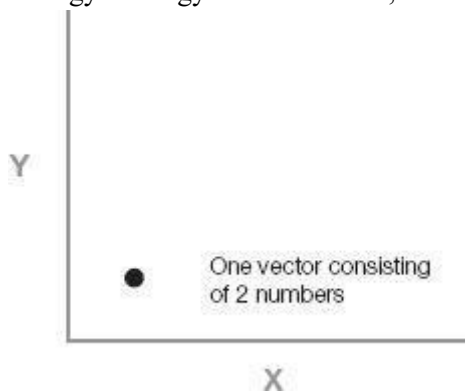
Bár nyilvánvaló volt, hogy az agy masszívan párhuzamos, számomra lehetetlennek tűnt, hogy különálló 3,000 hallási jeleken végezzen mintaillesztést. Kétkedtem abban, hogy az evolúció ennyire hatástalan lett volna. Ma már tudjuk, hogy a hallóidegben valóban nagyon jelentős adatesökkentés zajlik, mielőtt a hangjelek egyáltalán elérnék a neokortexet.

A szoftveralapú beszédfelismerőinkben szintén szoftveresen megvalósított szűrőket használtunk, pontosan tizenhatot (amit később harminckettőre növeltünk, mivel úgy találtuk, hogy ennél sokkal magasabbra emelésnek nem sok haszna van). A mi rendszerünkben tehát minden egyes időpontot tizenhat számmal reprezentáltunk. Ezt a tizenhat adatfolyamot kellett egybe redukálnunk, miközben ki kellett emelnünk a beszéd felismerésében jelentősnek számító jellemzőket.

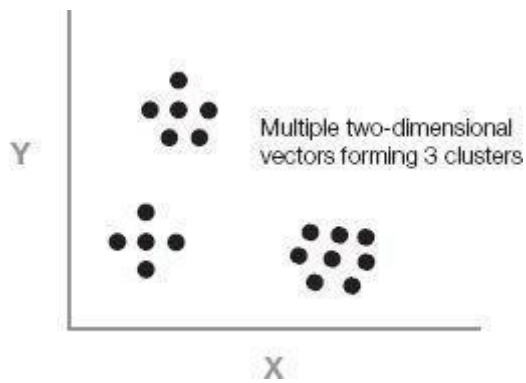
Ehhez egy matematikailag optimális technikát, az úgynevezett vektor kvantálást használtuk. Gondoljunk arra, hogy a hangot (legalábbis az egyik fülből) egy adott időpontban a szoftverünk tizenhat különböző számmal reprezentálta: ez a tizenhat frekvenciaszűrő kimenete. (Az emberi hallórendszerben ez a szám 3000 lenne, ami a 3000 cochlea belső szőrsejt kimenetét reprezentálja). A matematikai terminológiában minden ilyen számhalmazt (akár 3000 a biológiai esetben, akár a 16mi szoftveres megvalósításunkban) vektornak nevezünk.

Az egyszerűség kedvéért tekintsük a vektor kvantálás folyamatát két számból álló vektorokkal. Minden egyes vektor a kétdimenziós tér egy-egy pontjának tekinthető.

Ha ilyen vektorokból nagyon nagy mintánk van, és ábrázoljuk őket, akkor



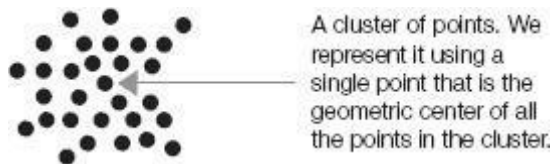
valószínűleg klaszterek kialakulását fogjuk észrevenni.



A klaszterek azonosításához el kell döntenünk, hogy hányat engedünk meg. A mi projektünkben általában 1024 klasztert engedélyeztünk, hogy meg tudjuk számozni őket, és minden klaszterhez 10 bites címkét rendelhessünk (mivel $2^{10} = 1024$). A vektorokból álló mintánk az általunk elvárt sokféleséget képviseli. Az első vektorokat 1,024 próbaképpen hozzárendeljük

hogy egy pontos klaszterek legyenek. Ezután megvizsgáljuk az 1025-ös vektort, és megkeressük azt a pontot, amelyikhez a legközelebb van. Ha ez a távolság nagyobb, mint az 1 024 pont bármelyik párja közötti legkisebb távolság, akkor azt egy új klaszter kezdetének tekintjük. Ezután az egymáshoz legközelebb eső két (egy pontos) klasztert egyetlen klaszterre zúzzuk össze. Így továbbra is 1024 klaszter marad. Az 1 025. vektor feldolgozása után az egyik ilyen klaszter már egynél több pontot tartalmaz. A pontokat továbbra is így dolgozzuk fel, mindig fenntartva az 1024 klasztert. Miután az összes pontot feldolgoztuk, minden egyes többpontos klasztert az adott klaszterben lévő pontok geometriai középpontjával ábrázolunk.

Ezt az iteratív folyamatot addig folytatjuk, amíg végig nem futunk az összes mintaponton. Általában több millió pontot dolgozunk fel ($1,024 \cdot 10^{10}$) klaszterbe; mi is használtunk ($2,048 \cdot 10^{11}$) vagy ($4,096 \cdot 10^{12}$) klasztert.



Minden klasztert egy vektor képvisel, amely az adott klaszterben lévő összes pont geometriai középpontjában van. Így a klaszter összes pontja és a klaszter középpontja közötti távolságok összege a lehető legkisebb.

Ennek a technikának az eredménye az, hogy a kiindulási több millió pont helyett (és a lehetséges pontok még nagyobb száma helyett) most mindössze 1024 pontra csökkentettük az adatokat, amelyek optimálisan használják ki a lehetőségek terét. A tér azon részei, amelyeket soha nem használunk, nem kapnak klasztereket.

Ezután minden klaszterhez hozzárendelünk egy számot (esetünkben 0-tól 1,023-ig). Ez a szám az adott klaszter redukált, "kvantált" reprezentációja, ezért nevezik a technikát vektor kvantálásnak. A jövőben érkező új bemeneti vektorokat ezután annak a klaszternek a száma reprezentálja, amelynek középpontja a legközelebb van ehhez az új bemeneti vektorhoz.

Most már előzetesen kiszámíthatunk egy táblázatot, amely tartalmazza minden klaszter középpontjának távolságát minden más középponttól. Így azonnal rendelkezésünkre áll ennek az új bemeneti vektornak (amelyet ezzel a kvantált ponttal - más szóval annak a klaszternek a számával, amelyhez ez az új pont a legközelebb van) a távolsága minden más klaszterhez. Mivel a pontokat csak a hozzájuk legközelebbi klaszterrel reprezentáljuk, most már tudjuk ennek a pontnak a távolságát minden más lehetséges ponttól, amelyik erre jön.

A fenti technikát csak két számot tartalmazó vektorok segítségével írtam le, de a tizenhat elemű vektorokkal való munka teljesen analóg az egyszerűbb példával. Mivel tizenhat különböző frekvenciasávot jelképező tizenhat számmal rendelkező vektorokat választottunk, rendszerünk minden egyes pontja egy pont volt a tizenhat dimenziós térben. Számunkra nehéz elképzelni egy háromnál (talán négynél, ha az időt is beleszámítjuk) több dimenziójú teret, de a matematikának nincsenek ilyen gátlásai.

Négy dolgot értünk el ezzel a folyamattal. Először is, jelentősen csökkentettük az adatok összetettségét. Másodsor, a tizenhat dimenziós adatokat egydimenziós adatokra csökkentettük (azaz minden minta egyetlen szám). Harmadszor, javítottuk az invariáns jellemzők megtalálásának képességét, mivel a lehetséges hangok terének azon részeit emeljük ki, amelyek a legtöbb információt közvetítik. A frekvenciák legtöbb kombinációja fizikailag lehetetlen vagy legalábbis nagyon valószínűtlen, így nincs okunk arra, hogy a valószínűtlen bemeneti kombinációknak ugyanolyan teret adjunk, mint a valószínű kombinációknak. Ez a technika az adatokat egyformán valószínű lehetőségekre csökkenti. A negyedik előnye, hogy egydimenziós mintafelismerőket használhatunk, még akkor is, ha az eredeti adatok sokkal több dimenzióból állnak. Ez bizonyult a leghatékonyabb megközelítésnek a rendelkezésre álló számítási erőforrások kihasználására.

Az elméd olvasása rejtett Markov-modellekkel

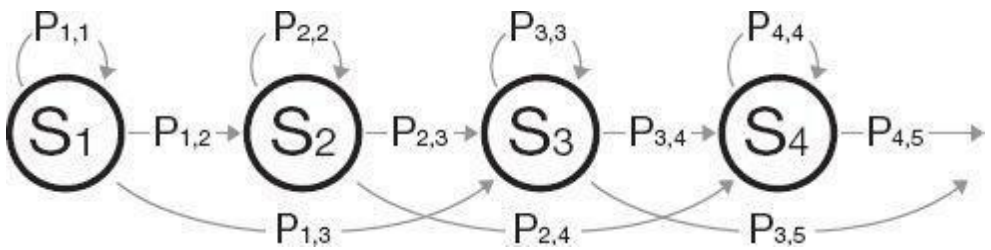
A vektorkvantálással úgy egyszerűsítettük az adatokat, hogy a kulcsfontosságú jellemzőket hangsúlyoztuk, de még mindig szükségünk volt egy olyan módra az invariáns jellemzők hierarchiájának ábrázolására, amely értelmet ad az új információknak. Mivel akkoriban (az 1980-as évek elején) húsz évig dolgoztam a mintafelismerés területén, tudtam, hogy az egydimenziós reprezentációk sokkal erősebbek, hatékonyabbak és alkalmasabbak az invariáns eredményekre. Az 1980-as évek elején még nem sokat tudtunk a neokortexről, de a különböző mintafelismerési problémákkal kapcsolatos tapasztalataim alapján feltételeztem, hogy az agy valószínűleg szintén egydimenziós reprezentációval redukálja a többdimenziós adatait (akár a szem, a fül, akár a bőr), különösen ahogy a fogalmak emelkednek a neokortex hierarchiájában.

A beszéd felismerési probléma esetében a beszédjelben lévő információ szerveződése a minták hierarchiájának tűnt, ahol minden egyes mintát egy lineáris elemsorozat képvisel, amelynek van egy előremenő iránya. Egy minta minden egyes eleme lehetett egy másik, alacsonyabb szintű minta, vagy a bemenet egy alapvető egysége (ami a beszéd felismerés esetében a mi kvantált vektoraink lennének).

Ezt a helyzetet a neokortex korábban bemutatott modelljével összhangban lévőnek fogja felismerni. Az emberi beszédet tehát az agyban lineáris minták hierarchiája hozza létre. Ha egyszerűen megvizsgálhatnánk ezeket a mintákat a beszélő személy agyában, akkor egyszerű lenne összevetni az új beszéd kifejezéseit az agyi mintákkal, és megérteni, hogy mit mond az illető. Sajnos azonban nincs közvetlen hozzáférésünk a beszélő agyához - az egyetlen információnk az, amit ténylegesen mondott. Természetesen ez a beszélt nyelv lényege - a beszélő megosztja az elméjének egy darabját a kijelentésével.

Ezért elgondolkodtam: Vajon létezik-e olyan matematikai technika, amely lehetővé teszi, hogy a beszélő agyában lévő mintákra következtessünk a beszélt szavai alapján? Egyetlen megszólalás nyilvánvalóan nem lenne elegendő, de ha nagyszámú mintával rendelkeznenénk, fel tudnánk-e használni ezt az információt arra, hogy lényegében leolvassuk a beszélő neokortexében lévő mintákat (vagy legalábbis megfogalmazzunk valami matematikailag egyenértékűt, ami lehetővé tenné számunkra, hogy felismerjük az új megszólalásokat)?

Az emberek gyakran nem tudják értékelni, hogy a matematika milyen hatalmas erejű tud lenni - ne feledjük, hogy az emberi tudás nagy részét a másodperc töredéke alatt tudjuk keresni a keresőmotorokkal, ami egy matematikai technikán alapul. A beszéd felismerési probléma esetében, amellyel az 1980-as évek elején szembesültem, kiderült, hogy a rejtett Markov-modellek technikája tökéletesen megfelel a célnak. Andrej Andrejevics Markov (1856-1922) orosz matematikus a hierarchikus állapotsorozatok matematikai elméletét építette fel. A modell alapja az volt, hogy az állapotokat egy láncban át lehet haladni, és ha ez sikerül, akkor a hierarchia következő, magasabb szintjén lévő állapotot lehet kiváltani. Ismerősen hangzik?



Egy egyszerű példa egy rejtett Markov-modell egy rétegére. S1-től S4-ig a "rejtett" belső állapotok. A $P_{i,j}$ átmenetek mindegyike a S_i állapotból az S_j állapotba való átlépés valószínűségét jelenti. Ezeket a valószínűségeket a rendszer a képzési adatokból (beleértve a tényleges használat során) tanuló rendszer határozza meg. Egy új szekvenciát (például egy új beszélt kifejezést) összevetünk ezekkel a valószínűségekkal, hogy meghatározzuk annak valószínűségét, hogy ez a modell hozta létre a szekvenciát.

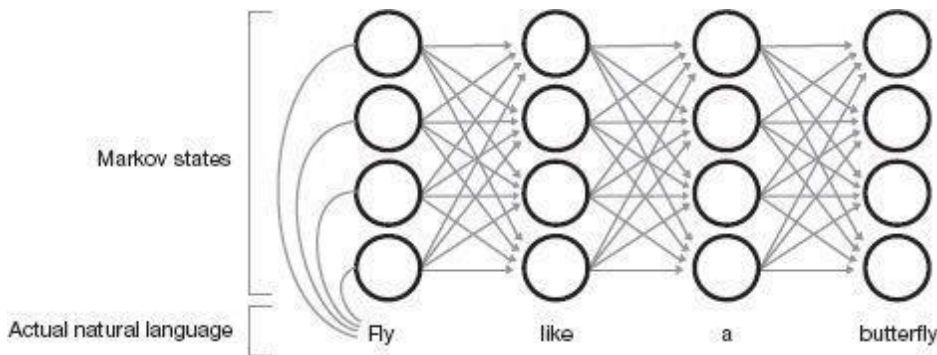
Markov modellje tartalmazta az egyes állapotok sikeres bekövetkezésének valószínűségeit. A továbbiakban olyan helyzetet tétélezett fel, amelyben egy rendszerben ilyen lineáris állapotsorozatok hierarchiája van, de ezeket nem lehet közvetlenül vizsgálni - innen a *rejtett* Markov-modellek elnevezés. A hierarchia legalsó szintje jeleket bocsát ki, és csak ezeket láthatjuk. Markov matematikai technikát biztosít annak kiszámítására, hogy a megfigyelt kimenet alapján milyen valószínűségeknél kell lenniük az egyes átmeneteknek. A módszert később Norbert Wiener 1923-ban finomította. Wiener finomítása módot adott a Markov-modellben lévő kapcsolatok meghatározására is; lényegében minden olyan kapcsolatot, amelynek valószínűsége túl alacsony, úgy tekintettek, hogy nem létezik. Az emberi neokortex lényegében így nyírja ki a kapcsolatokat - ha ritkán vagy soha nem használják őket, akkor valószínűtlennek tekintik és eltávolítják őket. Esetünkben a megfigyelt kimenet a beszélő személy által létrehozott beszédjel, a Markov-modell állapotvalószínűségei és kapcsolatai pedig az azt létrehozó neokortikális hierarchiát alkotják.

Olyan rendszert képzeltem el, amelyben emberi beszédmintákat veszünk, a rejtett Markov-modell technikát alkalmazzuk a kapcsolatokkal és valószínűségekkal rendelkező állapotok hierarchiájának levezetésére (lényegében egy szimulált neokortex a beszéd előállításához), majd ezt a levezetett hierarchikus állapothálózatot használjuk az új kifejezések felismerésére. Egy beszélőfüggetlen rendszer létrehozásához sok különböző személytől származó mintákat használnánk a rejtett Markov-modellek betanításához. Azáltal, hogy a hierarchiák elemét hozzáadtuk a nyelvben lévő információ hierarchikus jellegének ábrázolásához, ezeket megfelelően hierarchikus rejtett Markov-modelleknek (HHMM) neveztük el. A Kurzweil Applied Intelligence-nél dolgozó kollégáim szkeptikusak voltak azzal kapcsolatban, hogy ez a technika működni fog, mivel ez egy önszerveződő módszer, amely a neurális hálókra emlékeztet, amelyek kiestek a divatból, és amelyekkel kevés sikert értünk el. Rámutattam, hogy a neurális hálós rendszerben a hálózat fix, és nem alkalmazkodik a bemenethez: A súlyok alkalmazkodnak, de a kapcsolatok nem. A Markov-modell rendszerben, ha helyesen volt beállítva, a rendszer a nem használt kapcsolatokat úgy selejtezi, hogy lényegében a topológia alkalmazkodik.

Létrehoztam egy "skunk works" projektet (ez egy szervezeti kifejezés egy olyan projektre, amely a kitaposott ösvényen kívül van, és kevés hivatalos erőforrással rendelkezik), amely belőlem, egy részmunkaidős programozóból és egy villamosmérnökből állt (a frekvenciaszűrő bank létrehozására). Kollégáim meglepetésére erőfeszítésünk nagyon sikeresnek bizonyult, mivel sikerült nagy szókincset tartalmazó beszédet nagy pontossággal felismerni.

Ezt a kísérletet követően minden további beszédfelismerési kísérletünk hierarchikus rejtett Markov-modelleken alapult. Úgy tűnik, más beszédfelismerő cégek ettől függetlenül fedezték fel ennek a módszernek az értékét, és az 1980-as évek közepe óta a legtöbb automatizált beszédfelismerési munka ezen a megközelítésen alapul. A rejtett Markov-modelleket a beszédészintézisben is használják - ne feledjük, hogy biológiai agykérgi hierarchiánkat nemcsak a bemenet felismerésére, hanem a kimenet, például a beszéd és a fizikai mozgás előállítására is használjuk.

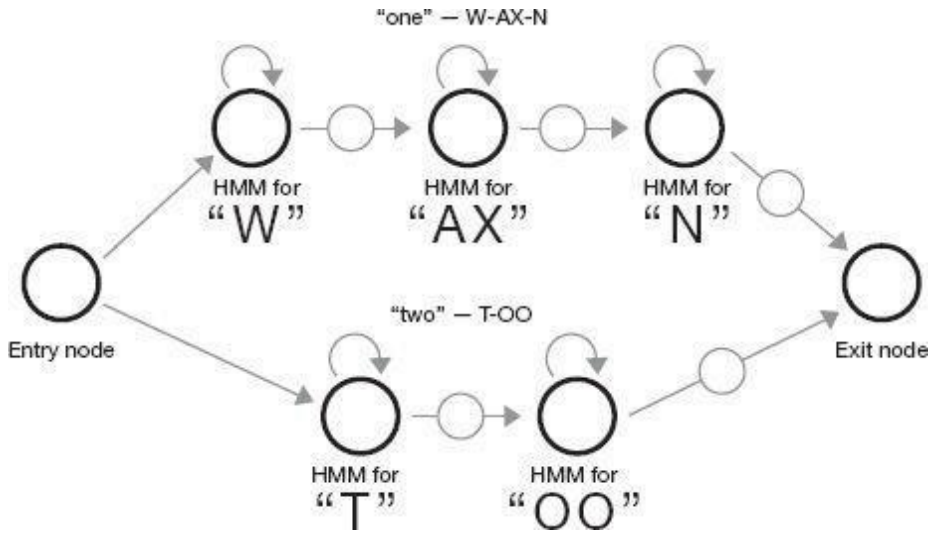
A HHMM-eket olyan rendszerekben is használják, amelyek megértik a természetes nyelvű mondatok jelentését, ami a fogalmi hierarchiában való felfelé haladást jelenti.



Rejtett Markov-állapotok és lehetséges átmenetek egy természetes nyelvű szövegben lévő szavak sorozatának előállításához.

Ahhoz, hogy megértsük, hogyan működik a HHMM-módszer, egy olyan hálózattal kezdünk, amely az összes lehetséges állapotátmenetből áll. A fent leírt vektor kvantálási módszer itt kritikus, mert különben túl sok lehetőség lenne, amit figyelembe kellene venni.

Íme egy lehetséges egyszerűsített kezdeti topológia:



Egy egyszerű rejtett Markov-modell topológia két beszélt szó felismerésére. A mintául szolgáló kijelentéseket egyenként dolgozzuk fel. Mindegyiknél iteratív módon módosítjuk az átmenetek valószínűségeit, hogy jobban tükrözzék az éppen feldolgozott bemeneti mintát. A beszédfelismerésben használt Markov-modellek kódolják annak valószínűségét, hogy az egyes fonémákban meghatározott hangminták találhatóak, hogy a fonémák hogyan befolyásolják egymást, és a fonémák valószínű sorrendjét. A rendszer a nyelvi struktúra magasabb szintjein is tartalmazhat valószínűségi hálózatokat, például a szavak sorrendjét, a mondatok felvételét, és így tovább a nyelvi hierarchiában.

Míg a korábbi beszédfelismerő rendszereink a fonémák szerkezetére és szekvenciáira vonatkozó, emberi nyelvészek által kifejezetten kódolt szabályokat építettek be, az új HHMM-alapú rendszernek nem mondták meg kifejezetten, hogy az angol nyelvben negyvennégy fonéma van, hogy az egyes fonémákhoz milyen vektorok szekvenciái valószínűek, vagy hogy milyen fonémák szekvenciái valószínűbbek, mint mások. Hagytuk, hogy a rendszer maga fedezze fel ezeket a "szabályokat" több ezer órányi átírt emberi beszédadatból. Ennek a megközelítésnek az az előnye a kézzel kódolt szabályokkal szemben, hogy a modellek olyan valószínűségi szabályokat fejlesztenek ki, amelyekkel az emberi szakértők gyakran nincsenek tisztában. Észrevettük, hogy a rendszer által az adatokból automatikusan megtanult szabályok közül sok finom, de fontos módon különbözik az emberi szakértők által felállított szabályoktól.

Miután a hálózatot betanítottuk, megpróbáltuk felismerni a beszédet úgy, hogy figyelembe vettük a hálózaton keresztülvezető alternatív utakat, és kiválasztottuk azt az utat, amely a legvalószínűbb volt a látott bemeneti vektorok tényleges sorrendje alapján. Más szóval, ha olyan állapotok sorozatát láttuk, amely valószínűsíthetően ezt a beszédet eredményezte, akkor arra következtettünk, hogy a beszéd ebből az agykérgi szekvenciából származik. Ez a szimulált HHMM-alapú neokortex szócímkéket is tartalmazott, így képes volt javaslatot tenni a hallottak átírására.

Ezután tovább tudtuk javítani az eredményeinket, ha folytattuk a hálózat edzését, miközben felismerésre használtuk. Amint azt már tárgyaltuk, a biológiai neokortikális hierarchiánk minden szintjén egyidejű felismerés és tanulás is zajlik.

Evolúciós (genetikai) algoritmusok

Van még egy fontos szempont: Hogyan állítsuk be azt a sok paramétert, amely egy mintafelismerő rendszer működését szabályozza? Ezek közé tartozhat a vektorok száma, amelyeket a vektor kvantálási lépésben megengedünk, a hierarchikus állapotok kezdeti topológiája (mielőtt a rejtett Markov-modell folyamatának képzési fázisa visszanyesné őket), a felismerési küszöb a hierarchia minden szintjén, a méretparaméterek kezelését szabályozó paraméterek és még sok más. Ezeket intuíciónk alapján is megállapíthatjuk, de az eredmények messze nem lesznek optimálisak.

Ezeket a paramétereket "isteni paramétereknek" nevezzük, mert a rejtett Markov-modellek topológiájának meghatározására szolgáló önszerveződő módszer előtt kerülnek beállításra (vagy biológiai esetben, mielőtt a személy megtanulja a leckéit azáltal, hogy hasonló módon kapcsolatokat hoz létre az agykérgi hierarchiában). Ez talán téves elnevezés, mivel ezeket a kezdeti DNS-alapú tervezési részleteket a biológiai evolúció határozza meg, bár egyesek Isten kezét látják ebben a folyamatban (és bár én az evolúciót spirituális folyamatnak tartom, ez a vita megfelelően a [9. fejezetbe](#) tartozik).

Amikor ezeket az "isteni paramétereket" kellett beállítani a szimulált hierarchikus tanulási és felismerési rendszerünkben, ismét a természetből vettük a példát, és úgy döntöttünk, hogy kialakítjuk őket - esetünkben az evolúció szimulációjával. Az úgynevezett genetikai vagy evolúciós algoritmusokat (GA) használtuk, amelyek szimulált szexuális reprodukciót és mutációkat tartalmaznak.

Az alábbiakban leegyszerűsítve ismertetjük, hogyan működik ez a módszer. Először is meghatározzuk egy adott probléma lehetséges megoldásainak kódolási módját. Ha a probléma egy áramkör tervezési paramétereinek optimalizálása, akkor meghatározzuk az áramkört jellemző összes paraméter listáját (az egyes paraméterekhez meghatározott számú bitet rendelve). Ezt a listát tekintjük a genetikai algoritmusban genetikai kódnak. Ezután véletlenszerűen több ezer vagy több genetikai kódot generálunk. Minden ilyen genetikai kódot (amely a tervezési paraméterek egy-egy készletét képviseli) szimulált "megoldási" organizmusnak tekintünk.

Most minden egyes szimulált szervezetet egy szimulált környezetben értékelünk egy meghatározott módszerrel az egyes paraméterek értékelésére. Ez az értékelés a genetikai algoritmus sikerének kulcsa. Példánkban lefuttatnánk minden egyes, a paraméterek által generált programot, és megfelelő kritériumok alapján értékelnénk (elvégezte-e a feladatot, mennyi ideig tartott, stb.). A legjobb megoldású organizmusok (a legjobb tervek) maradhatnak életben, a többiek pedig kiesnek.

Most a túlélők mindegyike addig szaporodik, amíg el nem éri az azonos számú

megoldási lényt. Ezt a szexuális szaporodás szimulálásával végezzük: Más szóval, új utódokat hozunk létre, ahol minden új lény genetikai kódjának egy részét az egyik szülőjétől, egy másik részét pedig egy másik szülőtől kapja. Általában nem teszünk különbséget hím vagy nőstény élőlények között; elég, ha két tetszőleges szülőtől bármelyik utódot létrehozunk, tehát itt lényegében azonos neműek házasságáról beszélünk. Ez talán nem olyan érdekes, mint a természetben a szexuális szaporodás, de a lényeg itt a két szülő megléte. Ahogy ezek a szimulált organizmusok szaporodnak, megengedünk némi mutációt (véletlenszerű változást) a kromoszómákban.

A szimulált evolúció egy generációját már meghatároztuk; most megismételjük ezeket a lépéseket minden további generáció esetében. Minden generáció végén meghatározzuk, hogy mennyit javultak a tervek (azaz kiszámítjuk az értékelési függvény átlagos javulását az összes túlélő organizmusra vonatkozóan). Amikor a tervezői lények értékelésének egyik generációról a másikra történő javulásának mértéke nagyon kicsi lesz, leállítjuk ezt az iteratív ciklust, és az utolsó generáció legjobb tervét (terveit) használjuk. (A genetikai algoritmusok algoritmikus leírását lásd ebben a végjegyzetben.)¹¹

A genetikai algoritmus kulcsa az, hogy az emberi tervezők nem közvetlenül programoznak be egy megoldást; inkább hagyjuk, hogy a megoldás a szimulált verseny és a fejlesztés iteratív folyamatán keresztül alakuljon ki. A biológiai evolúció okos, de lassú, ezért az intelligencia fokozása érdekében jelentősen felgyorsítjuk a nehézkes tempót. A számítógépek elég gyorsak ahhoz, hogy sok generációt szimuláljanak órák vagy napok alatt, és alkalmanként akár hetekig is futtattuk őket, hogy több százezer generációt szimuláljanak. De ezen az iteratív folyamaton csak egyszer kell végigmennünk; amint hagyjuk, hogy ez a szimulált evolúció végigfusson, a kialakult és nagyon kifinomult szabályokat gyorsan alkalmazhatjuk valós problémákra. Beszédfelismerő rendszereink esetében a hálózat kezdeti topológiájának és más kritikus paraméterek evolúciójára használtuk őket. Két önszerveződő módszert használtunk tehát: egy GA-t a biológiai evolúció szimulálására, amely egy adott agykérgi dizájnt eredményezett, és HHMM-eket az emberi tanulást kísérő agykérgi szerveződés szimulálására.

A GA sikerének másik fontos feltétele az egyes lehetséges megoldások értékelésének érvényes módszere. Ezt az értékelést gyorsan kell elvégezni, mivel a szimulált evolúció minden egyes generációjánál több ezer lehetséges megoldást kell figyelembe venni. A GA-k jól kezelik a túl sok változót tartalmazó problémákat, amelyek esetében nem lehet pontos analitikus megoldásokat kiszámítani. Egy hajtómű tervezése például több mint száz változót tartalmazhat, és több tucatnyi megkötésnek kell megfelelni; a General Electric kutatói által használt GA-k képesek voltak olyan sugárhajtóművek tervezésére, amelyek a hagyományos módszereknél pontosabban megfeleltek a megkötéseknek.

A GA-k használatakor azonban óvatosnak kell lennie, hogy mit kér. Egy genetikus algoritmust használtak egy tömbhalmazási probléma megoldására, és tökéletes megoldással állt elő... kivéve, hogy több ezer lépésből állt. Az emberi programozók elfelejtették a lépések számának minimalizálását az értékelő függvényükbe foglalni.

Scott Drave Electric Sheep projektje egy GA, amely művészetet termel. Az értékelési

funkció emberi értékelőket használ, egy nyílt forráskódú, több ezer ember részvételével zajló együttműködésben. A művészet az időben mozog, és megtekinthető az electricsheep.org oldalon.

A beszédfelismerés esetében a genetikai algoritmusok és a rejtett Markov-modellek kombinációja rendkívül jól működött. Az evolúció GA-val történő szimulálása jelentősen javítani tudta a HHMM-hálózatok teljesítményét. Amire az evolúció jutott, az messze felülmúlta az eredeti tervünket, amely az intuíciónkon alapult.

Ezután kísérleteztünk egy sor kisebb változtatás bevezetésével a teljes rendszerben. Például perturbációkat (kisebb véletlenszerű változtatásokat) végeztünk a bemeneten. Egy másik ilyen változtatás az volt, hogy a szomszédos Markov-modellek "szivárogtak" egymásba azáltal, hogy az egyik Markov-modell eredményei befolyásolják a "közeli" modelleket. Bár akkor még nem vettük észre, a módosítások fajtái, amelyekkel kísérleteztünk, nagyon hasonlítanak a biológiai agykérgi struktúrákban előforduló módosítások típusaihoz.

Eleinte az ilyen változások rontják a teljesítményt (a felismerés pontosságával mérve). Ha azonban az evolúciót (azaz a GA-t) újraindítanánk ezekkel a módosításokkal, a rendszer ennek megfelelően alkalmazkodna, optimalizálva azt a bevezetett módosításokra. Általában ez helyreállítaná a teljesítményt. Ha ezután eltávolítanánk a bevezetett módosításokat, a teljesítmény ismét romlana, mivel a rendszer úgy lett evolválva, hogy kompenzálja a módosításokat. Az adaptált rendszer függővé vált a változásoktól.

Az egyik olyan változtatás, amely valóban segített a teljesítményen (a GA újrafuttatása után), az volt, hogy kis véletlenszerű változásokat vezettünk be a bemenetbe. Ennek oka az önszerveződő rendszereknél jól ismert "túlillesztési" probléma. Fennáll a veszélye annak, hogy egy ilyen rendszer túl általánosít a képzési mintában szereplő konkrét példákra. A bemenet véletlenszerű módosításával az adatokban lévő invariánsabb minták megmaradnak, és a rendszer ezáltal megtanulja ezeket a mélyebb mintákat. Ez csak akkor segített, ha a GA-t a randomizációs funkció bekapcsolásával futtattuk újra.

Ez dilemmát okoz a biológiai agykérgi áramkörök megértésében. Észrevettük például, hogy a biológiai kapcsolatok kialakulásának módjából adódóan valóban lehet egy kis szivárgás az egyik agykérgi kapcsolatból a másikba: Az axonok és dendritek elektrokémiája nyilvánvalóan ki van téve a közeli kapcsolatok elektromágneses hatásainak. Tegyük fel, hogy képesek vagyunk egy olyan kísérletet lefuttatni, ahol ezt a hatást egy valódi agyban eltávolítjuk. Ezt nehéz lenne ténylegesen végrehajtani, de nem feltétlenül lehetetlen. Tegyük fel, hogy elvégezzünk egy ilyen kísérletet, és azt találjuk, hogy az agykérgi áramkörök kevésbé hatékonyan működnek ezen idegszivárgás nélkül. Akkor arra a következtetésre juthatnánk, hogy ez a jelenség az evolúció nagyon okos tervezése volt, és döntő fontosságú volt ahhoz, hogy az agykéreg elérje a teljesítményszintjét. Rámutathatnánk továbbá, hogy egy ilyen eredmény azt mutatja, hogy a minták fogalmi hierarchiában felfelé irányuló áramlásának és a jóslatok hierarchiában lefelé irányuló áramlásának rendezett modellje valójában sokkal bonyolultabb volt a kapcsolatoknak ez a bonyolult egymásra hatása miatt.

De ez nem feltétlenül lenne pontos következtetés. Tekintsük a HHMM-eken alapuló szimulált kéreggel kapcsolatos tapasztalatainkat, amelyben egy, az interneuronális keresztbeszélgetéshez nagyon hasonló módosítást valósítottunk meg. Ha ezután lefuttattuk az evolúciót ezzel a jelenséggel a helyén, a teljesítmény helyreállt (mert az evolúciós

folymat alkalmazodott hozzá). Ha ezután eltávolítottuk a keresztbeszélgetést, a teljesítmény ismét romlott. A biológiai esetben az evolúciót (vagyis a biológiai evolúciót) valóban úgy "futtatták", hogy ez a jelenség megmaradt. A rendszer részletes paramétereit tehát a biológiai evolúció úgy állította be, hogy azok függjenek ezektől a tényezőktől, így ezek megváltoztatása negatívan befolyásolja a teljesítményt, hacsak nem futtatjuk újra az evolúciót. Ez megvalósítható a szimulált világban, ahol az evolúció csak napokig vagy hetekig tart, de a biológiai világban ehhez több tízezer évre lenne szükség.

Hogyan tudjuk tehát megmondani, hogy a biológiai neokortex egy bizonyos tervezési jellemzője a biológiai evolúció által bevezetett létfontosságú újítás - vagyis az intelligenciánk szintjéhez szükséges - vagy csupán egy olyan műtárgy, amelytől a rendszer tervezése most függ, de anélkül is fejlődhetett volna? Erre a kérdésre egyszerűen úgy tudunk választ adni, hogy szimulált evolúciót futtatunk le a tervezés részleteinek ezen különleges variációival és azok nélkül (például a kapcsolatok keresztbeszélgetésével és anélkül). Ezt még a biológiai evolúcióval is megtehetjük, ha egy mikroorganizmus-kolónia evolúcióját vizsgáljuk, ahol a generációkat órákban mérjük, de ez nem praktikus az olyan összetett organizmusok esetében, mint az ember. Ez egy újabb a biológia számos hátránya közül.

Visszatérve a beszédfelismeréssel kapcsolatos munkánkhoz, azt találtuk, hogy ha az evolúciót (azaz a GA-t) *külön* futtatjuk a kezdeti terven a

(1) a fonémák belső szerkezetét modellező hierarchikus rejtett Markov-modelleket és (2) a szavak és mondatok szerkezetét modellező HHMM-eket, még jobb eredményeket kaptunk. A rendszer mindkét szintje HHMM-eket használt, de a GA tervezési variációkat fejlesztett ki e különböző szintek között. Ez a megközelítés még mindig lehetővé tette a két szint között előforduló jelenségek modellezését, például a fonémák elkenődését, ami gyakran előfordul, amikor bizonyos szavakat egymás után fűzünk (például a "Hogy vagytok?" mondatból "Hogy vagytok?" lehet "Hogy vagytok?").

Valószínű, hogy hasonló jelenség játszódott le a különböző biológiai agykérgi régiókban is, amelyekben az általuk feldolgozott mintázatok típusai alapján alakultak ki apró különbségek. Míg ezek a régiók mind ugyanazt az alapvető neokortikális algoritmust használják, a biológiai evolúciónak elég ideje volt arra, hogy az egyes régiók kialakítását úgy finomhangolja, hogy az optimális legyen az adott mintához. Azonban, ahogyan azt korábban tárgyaltam, az idegtudósok és neurológusok jelentős plaszticitást észleltek ezeken a területeken, ami alátámasztja az általános neokortikális algoritmus gondolatát. Ha az egyes régiók alapvető módszerei gyökeresen eltérőek lennének, akkor az agykérgi régiók közötti ilyen felcserélhetőség nem lenne lehetséges.

A kutatásunk során az önszerveződő módszerek e kombinációjával létrehozott rendszerek nagyon sikeresek voltak. A beszédfelismerésben először voltak képesek teljesen folyamatos beszédet és viszonylag korlátlan szókincset kezelni. Magas pontossági arányt tudtunk elérni a legkülönbözőbb beszélők, akcentusok és dialektusok esetében. E könyv írásakor a technika jelenlegi állását a Nuance (korábban Kurzweil Computer Products) cég Dragon Naturally Speaking (11.5. verzió) nevű terméke képviseli PC-re. Azt javaslom, hogy az emberek próbálják ki, ha szkeptikusak a kortárs beszédfelismerés teljesítményével kapcsolatban - a pontosság gyakran 99 százalékos vagy annál magasabb, miután néhány percig gyakoroltuk a hangunkat folyamatos beszéd és viszonylag korlátlan szókincs esetén.

A Dragon Dictation egy egyszerűbb, de még mindig lenyűgöző ingyenes alkalmazás az iPhone-ra, amely nem igényel hangképzést. A Siri, a mai Apple iPhone-ok személyi asszisztense ugyanazt a beszédfelismerő technológiát használja, a természetes nyelvi megértés kezelésére szolgáló bővítésekkel.

Ezeknek a rendszereknek a teljesítménye a matematika erejéről tanúskodik. Lényegében azt számoljuk ki, hogy mi zajlik a beszélő neokortexében - még akkor is, ha nincs közvetlen hozzáférésünk az adott személy agyához -, ami elengedhetetlen lépés ahhoz, hogy felismerjük, mit mond az illető, és az olyan rendszerek esetében, mint a Siri, mit jelentenek ezek a kijelentések. Elgondolkodhatunk azon, hogy ha valóban belenézünk a beszélő neokortexébe, vajon látnánk-e a szoftver által kiszámított hierarchikus rejtett Markov-modelleknek megfelelő kapcsolatokat és súlyokat? Szinte biztos, hogy nem találnánk pontos egyezést; az idegsejtek struktúrái minden bizonnyal sok részletben eltérnének a számítógépben lévő modellektől. Én azonban azt állítom, hogy a tényleges biológia és az azt utánzó kísérletünk között alapvető matematikai egyenértékűségnek kell lennie, méghozzá nagyfokú pontossággal; különben ezek a rendszerek nem működnének olyan jól, mint ahogyan működnek.

LISP

A LISP (LISt Processor) egy számítógépes nyelv, amelyet eredetileg John McCarthy (1927-2011), a mesterséges intelligencia úttörője írt le 1958-ban. Ahogy a neve is mutatja, a LISP listákkal foglalkozik. Minden LISP-kijelentés egy elemekből álló lista; minden elem vagy egy másik lista, vagy egy "atom", amely egy számot vagy egy szimbólumot alkotó, nem redukálható elem. Egy listában szereplő lista lehet maga a lista, ezért a LISP képes a rekurzíóra. A LISP utasítások egy másik módja, hogy rekurzívak lehetnek, ha egy lista tartalmaz egy listát, és így tovább, amíg az eredeti lista meg nem adódik. Mivel a listák tartalmazhatnak listákat, a LISP hierarchikus feldolgozásra is képes. Egy lista lehet olyan feltételes, hogy csak akkor "tüzel", ha az elemei teljesülnek. Ily módon az ilyen feltételes hierarchiák felhasználhatók egy minta egyre absztraktabb tulajdonságainak azonosítására.

A LISP az 1970-es években és az 1980-as évek elején lett a mesterséges intelligencia közösségének divatja. A korábbi évtized LISP-rajongóinak az volt a hiedelme, hogy a nyelv az emberi agy működését tükrözi, azaz bármilyen intelligens folyamatot a legkönnyebben és leghatékonyabban LISP-ben lehet kódolni. Ezt követően a LISP-értelmezőket és kapcsolódó LISP-termékeket kínáló "mesterséges intelligencia" cégek mini-boomja következett, de amikor az 1980-as évek közepén nyilvánvalóvá vált, hogy a LISP önmagában nem jelent rövidítést az intelligens folyamatok létrehozásához, a befektetési lufi összeomlott.

Kiderült, hogy a LISP-rajongók nem tévedtek teljesen. Lényegében minden egyes mintafelismerő a neokortexben egy LISP-kijelentésnek tekinthető - mindegyik egy elemekből álló listát alkot, és minden elem lehet egy másik lista. A neokortex tehát valóban olyan szimbolikus jellegű listafeldolgozással foglalkozik, amely nagyon hasonló ahhoz, ami egy LISP-programban zajlik. Ráadásul mind a 300 millió LISP-szerű "utasítást" egyszerre

dolgozza fel. A LISP világából azonban két fontos tulajdonság hiányzott, az egyik a tanulás. A LISP programokat emberi programozóknak kellett sorról sorra kódolniuk. Voltak ugyan kísérletek a LISP programok automatikus kódolására különböző módszerekkel, de ezek nem képezték a nyelv koncepciójának szerves részét. A neokortex ezzel szemben önmagát programozza, "utasításait" (vagyis a listákat) saját tapasztalataiból és saját visszacsatolási hurkaiból származó értelmes és cselekvésre alkalmas információkkal tölti meg. Ez a neokortex működésének egyik legfontosabb alapelve: Minden egyes mintafelismerője (vagyis minden egyes LISP-szerű utasítás) képes kitölteni a saját listáját, és mind felfelé, mind lefelé összekapcsolni magát más listákkal. A második különbség a méretparaméterek. Létre lehetne hozni a LISP egy olyan (LISP-ben kódolt) változatát, amely lehetővé tenné az ilyen paraméterek kezelését, de ezek nem részei az alapnyelvnek.

A LISP összhangban van a mesterséges intelligencia eredeti filozófiájával, amely a problémák intelligens megoldásának megtalálását és közvetlen számítógépes nyelveken történő kódolását tűzte ki célul. Az első kísérlet egy olyan önszerveződő módszerre, amely a tapasztalatok alapján tanítja magát - a neurális hálók - nem volt sikeres, mivel nem biztosított eszközt a rendszer topológiájának a tanulás hatására történő módosítására. A hierarchikus rejtett Markov-modell ezt hatékonyan biztosította a metszési mechanizmusa révén. Ma a HHMM matematikai rokonaival együtt a mesterséges intelligencia világának jelentős részét teszi ki.

A LISP és a neokortex listaszervezetének hasonlóságára vonatkozó megfigyelés következménye az az érv, amelyet azok hoznak fel, akik ragaszkodnak ahhoz, hogy az agy túl bonyolult ahhoz, hogy megértsük. Ezek a kritikusok rámutatnak arra, hogy az agyban trilliónyi kapcsolat van, és mivel mindegyiknek kifejezetten a tervezés miatt kell ott lennie, ezek a kapcsolatok trilliónyi kódsorral egyenértékűek. Mint láttuk, becsléseim szerint a neokortexben körülbelül 300 millió mintafeldolgozó processzor található - vagy 300 millió lista, ahol a lista minden egyes eleme egy másik listára mutat (vagy a legalacsonyabb fogalmi szinten egy, a neokortexen kívüli, irreducibilis alapmintára). De 300 millió még mindig meglehetősen nagy számú LISP-kijelentés, és valóban nagyobb, mint bármely létező, ember által írt program.

Nem szabad azonban elfelejtenünk, hogy ezek a listák valójában nincsenek meghatározva az idegrendszer eredeti tervezésekor. Az agy maga hozza létre ezeket a listákat, és saját tapasztalatai alapján automatikusan összekapcsolja a szinteket. Ez a neokortex kulcsfontosságú titka. Azok a folyamatok, amelyek ezt az önszerveződést megvalósítják, sokkal egyszerűbbek, mint a neokortex kapacitását képező milliányi 300 kijelentés. Ezek a folyamatok a genomban vannak meghatározva. Amint azt a [fejezetben 11](#) bemutatom, a genomban lévő egyedi információ mennyisége (veszteségmentes tömörítés után) az agyra alkalmazva körülbelül 25 millió bájttal, ami kevesebb mint egymillió kódsornak felel meg. A tényleges algoritmikus komplexitás még ennél is kevesebb, mivel a genetikai információ millió 25 bájttal nagy része a neuronok biológiai szükségleteihez tartozik, és nem kifejezetten az információfeldolgozó képességükhöz. Azonban még 25 millió bájtnyi tervezési információ is olyan komplexitási szint, amelyet kezelni tudunk.

Amint azt a [3. fejezetben](#) tárgyaltam, Jeff Hawkins és Dileep George 2003-ban és 2004-ben kidolgozta a neokortex hierarchikus listákat tartalmazó modelljét, amelyet Hawkins és Blakeslee 2004-ben megjelent *On Intelligence című* könyvében ismertetett. A hierarchikus időbeli memóriamódszer naprakészebb és nagyon elegáns bemutatása a következőkben olvasható

Dileep George 2008-as doktori disszertációjában található.¹² A Numenta a NuPIC (Numenta Platform for Intelligent Computing) nevű rendszerben valósította meg, és olyan ügyfelek számára fejlesztett mintafelismerő és intelligens adatbányászati rendszereket, mint a Forbes és a Power Analytics Corporation. Miután a Numentánál dolgozott, George új céget alapított Vicarious Systems néven, a Founder Fund (amelyet Peter Thiel, a Facebook mögött álló kockázati tőkés és Sean Parker, a Facebook első elnöke irányít) és a Dustin Moskovitz, a Facebook társalapítója által vezetett Good Ventures finanszírozásával. George jelentős előrelépésről számol be a jelentős számú hierarchiával rendelkező információk automatikus modellezése, tanulása és felismerése terén. Rendszerét "rekurzív agykérgi hálózatnak" nevezi, és alkalmazásokat tervez többek között az orvosi képzés és a robotika területén. A hierarchikus rejtett Markov-modellek technikája matematikailag nagyon hasonlít ezekre a hierarchikus memóriarendszerekre, különösen, ha megengedjük, hogy a HHMM-rendszer maga szervezze meg a mintafelismerő modulok közötti kapcsolatait. Mint korábban említettük, a HHMM-ek egy további fontos elemet biztosítanak, amely az egyes bemenetek nagyságának várható eloszlásának (valamilyen kontinuumon) modellezése a vizsgált minta létezésének valószínűségének kiszámításakor. Nemrégiben alapítottam egy új céget Patterns, Inc. néven, amely hierarchikus önszerveződő neokortikális modelleket kíván kifejleszteni, amelyek a HHMM-eket és a kapcsolódó technikákat használják a természetes nyelv megértése céljából. Fontos hangsúlyt fektetünk arra, hogy a rendszer a biológiai neokortexhez hasonlóan képes legyen saját hierarchiákat kialakítani. Az általunk elképzelt rendszer folyamatosan olvasni fogja az anyagok széles skáláját, például a Wikipédiát és más tudásforrásokat, valamint mindent meghallgat, amit mond, és mindent figyel, amit ír (ha hagyja). A cél az, hogy hasznos barátunkká váljon, aki válaszol a kérdéseinkre - még mielőtt megfogalmaznánk azokat -, és hasznos információkkal és tippekkel lát el bennünket a nap folyamán.

A mesterséges intelligencia mozgó határa: a kompetenciahierarchia megmászása

1. Egy hosszú, fárasztó beszéd, amelyet egy habos pitefelfújt mondott.
2. Egy gyermek által viselt ruhadarab, talán egy operahajó fedélzetén.
3. Tizenkét éven át tartó, Hrothgar király harcosainak felfalása miatt körözik; Beowulf tisztet bízták meg az üggyel.
4. Jelentheti azt, hogy fokozatosan fejlődik az elmében, vagy azt, hogy a

terhesség alatt hordoz.

5. Nemzeti Tanárnap és Kentucky Derby Nap.
6. Wordsworth azt mondta, hogy szárnyalnak, de soha nem kóborolnak.
7. Négybetűs szó, amely a ló patáján vagy a kaszinóban lévő kártyaosztó dobozon lévő vasszerelvényt jelöli.
8. Egy Verdi-opera 1846 harmadik felvonásában Isten eme ostorát szerelme, Odabella halálra késeli.

-Példák a *Jeopardy!* kérdésekre, amelyek mindegyikét Watson helyesen adta meg. A válaszok: habcsókos harangszó, pinafore, Grendel, gestate, May, skylark, shoe. A nyolcadik kérdésre Watson azt válaszolta: "Mi az az Attila?". A műsorvezető erre azt válaszolta: "Legyen pontosabb?". Watson így pontosított: "Mi Attila, a hun?", ami helyes.

A számítógép technikája a *Jeopardy!* feladványok megfejtésére pont úgy hangzott, mint az enyém. A gép a feladvány kulcsszavaira összpontosít, majd átfésüli a memóriáját (Watson esetében az emberi tudás 15 terabájtos adatbankját) asszociációk csoportjai után kutatva...

ezekkel a szavakkal. A legjobb találatokat szigorúan ellenőrzi az összes lehetséges kontextuális információval: a kategória neve; a keresett válasz típusa; a nyomban szereplő idő, hely és nem; és így tovább. És amikor eléggé biztosnak érzi magát, úgy dönt, hogy berreg. Mindez egy emberi *Jeopardy!* játékos számára azonnali, intuitív folyamat, de meg voltam győződve arról, hogy a motorháztető alatt az én agyam nagyjából ugyanezt csinálja.

-Ken Jennings, az emberi *Jeopardy!* bajnoka, aki kikapott Watsontól.

Én a magam részéről üdvözlöm új robot urainkat.

-Ken Jennings, a *Simpson család*ot idézve, miután kikapott Watsontól

Ó, Istenem. [Watson] intelligensebb, mint egy átlagos *Jeopardy!* játékos a *Jeopardy!* kérdések megválaszolásában. Ez lenyűgözően intelligens.

-Sebastian Thrun, a Stanford AI Lab korábbi igazgatója

Watson semmit sem ért. Ez egy nagyobb gőzhenger.

-Noam Chomsky

A mesterséges intelligencia mindenütt ott van körülöttünk - már nem a mi kezünk van a dugón. Az egyszerű aktus, hogy valakivel szöveges üzenetben, e-mailben vagy mobiltelefonhíváson keresztül kapcsolatba lépünk, intelligens algoritmusokat használ az információ továbbítására. Szinte minden termék, amelyet megérintünk, eredetileg az emberi és a mesterséges intelligencia együttműködésével kerül megtervezésre, majd automatizált gyárakban készül. Ha holnap az összes mesterséges intelligencia rendszer úgy döntene, hogy sztrájkba lép, civilizációnk megbénulna: Nem kaphatnánk pénzt a bankunkból, sőt, a pénzünk is eltűnne; a kommunikáció, a közlekedés és a gyártás is leállna. Szerencsére intelligens gépeink még nem elég intelligensek egy ilyen összeesküvés megszervezéséhez.

Ami ma újdonság a mesterséges intelligenciában, az a nyilvánosan elérhető példák látványos jellege. Gondoljunk például a Google önvezető autóira (amelyek e sorok írásakor már több mint 200 000 mérföldet tettek meg városokban és településeken), egy olyan technológiára, amely jelentősen kevesebb balesetet fog okozni, növeli az utak kapacitását, enyhíti az embereknek a vezetés feladatának elvégzésére vonatkozó igényét, és számos más előnnyel jár. A vezető nélküli autók bizonyos korlátozásokkal már most is legálisan közlekedhetnek a közutakon Nevadában, bár a széleskörű használat a világon csak az évtized végén várható. Az autókba már most is olyan technológiát építenek be, amely intelligensen figyeli az utat és figyelmezteti a vezetőt a közelgő veszélyekre. Az egyik ilyen technológia részben az MIT munkatársa, Tomaso Poggio által az agyban történő vizuális feldolgozás sikeres modelljén alapul. A MobilEye nevű rendszert Amnon Shashua, Poggio

korábbi posztdoktori hallgatója fejlesztette ki. Képes figyelmeztetni a vezetőt olyan veszélyekre, mint egy közelgő ütközés vagy egy autó elé szaladó gyermek, és nemrégiben olyan gyártók építették be az autókba, mint a Volvo és a BMW.

A könyv ezen részében több okból is a nyelvi technológiákra fogok összpontosítani. Nem meglepő, hogy a nyelv hierarchikus természete szorosan tükrözi gondolkodásunk hierarchikus természetét. A beszélt nyelv volt az első technológiánk, az írott nyelv a második. Saját munkám a mesterséges intelligencia területén, ahogy ez a fejezet is mutatja, nagymértékben a nyelvre összpontosított. Végül, a nyelv elsajátítása erőteljesen kihasználható képesség. Watson már több százmillió weboldalt olvasott el, és elsajátította az ezekben a dokumentumokban található tudást. Végző soron a gépek képesek lesznek elsajátítani a Web összes tudását - ami lényegében az ember-gép civilizációnk összes tudását jelenti.

Alan Turing (1912-1954) angol matematikus a névadó tesztjét arra alapozta, hogy egy számítógép képes-e természetes nyelven, szöveges üzenetek segítségével beszélgetni.¹³ Turing úgy vélte, hogy az emberi intelligencia egésze a nyelvben testesül meg és jelenik meg, és egyetlen gép sem tudna egyszerű nyelvi trükkökkel átmenni a Turing-teszten. Bár a Turing-teszt egy írott nyelvvel kapcsolatos játék, Turing úgy vélte, hogy a számítógép csak akkor tudna megfelelni rajta, ha valóban rendelkezne az emberi intelligencia szintjének megfelelő intelligenciával. A kritikusok azt javasolták, hogy az emberi szintű intelligencia valódi tesztjének a vizuális és auditív információk elsajátítását is tartalmaznia kellene.¹⁴ Mivel a saját mesterséges intelligencia projektjeim közül számos olyan érzékszervi információk elsajátítására tanítom a számítógépeket, mint az emberi beszéd, a betűformák és a zenei hangok, elvárható lenne, hogy támogassam az ilyen típusú információk bevonását egy valódi intelligenciatesztbe. Mégis egyetértek Turing eredeti felismerésével, miszerint a Turing-teszt csak szöveges változata elegendő. A vizuális vagy auditív input vagy output hozzáadása a teszthez valójában nem nehezítené meg a teszt teljesítését.

Nem kell ahhoz mesterséges intelligencia szakértőnek lenni, hogy az ember meghatódjon Watson teljesítményétől a *Jeopardy!* Bár eléggé értem a módszertant, amelyet számos kulcsfontosságú alrendszerében alkalmaznak, ez nem csökkenti az érzelmi reakciómat, amikor néztem - őt? - teljesítményét. Még ha tökéletesen értenék is, hogyan működik az összes rendszerem - amivel valójában senki sem rendelkezik -, akkor sem tudnánk megjósolni, hogyan reagálna Watson egy adott helyzetben. Több száz egymással kölcsönhatásban lévő alrendszerből áll, és ezek mindegyike egyszerre több millió konkurens hipotézist mérlegel, így a végeredményt lehetetlen megjósolni. Egy embernek évszázadokba telne, ha alaposan elemeznie kellene - utólag - Watson mérlegelését egyetlen három másodperces lekérdezés esetén.

Hogy folytassam a saját történetemet, az 1980-as évek végén és az 1990-es években kezdtünk el dolgozni a természetes nyelvi megértésen korlátozott területeken. Az egyik Kurzweil Voice nevű termékünkkel bármit meg lehetett beszélni, amiről csak akartunk, feltéve, hogy az a dokumentumok szerkesztésével kapcsolatos volt. (Például: "Az előző oldal harmadik bekezdését tegye ide.") Ez a korlátozott, de hasznos területen elég jól működött. Létrehoztunk orvosi szakterületi tudással rendelkező rendszereket is, hogy az

orvosok diktálhassák a betegjelentéseket. A rendszer elég ismerettel rendelkezett az olyan területekről, mint a radiológia és a patológia, ahhoz, hogy megkérdezze az orvost, ha a jelentésben valami nem tűnt egyértelműnek, és végigvezette az orvost a jelentéskészítés folyamatán. Ezek az orvosi jelentéskészítő rendszerek a Nuance milliárd dolláros üzletágává fejlődtek.

A természetes nyelv megértése, különösen az automatikus beszédfelismerés kiterjesztéseként, mára már a főáramba került. E könyv írásakor a Siri, az iPhone 4S automatikus személyi asszisztense nagy port kavart a mobil számítástechnika világában. A Siritől nagyjából bármit megkérhetünk, amire egy magára valamit is adó okostelefonnak képesnek kell lennie (például: "Hol lehet errefelé indiai ételt kapni?", "Küldj üzenetet a feleségemnek, hogy úton vagyok" vagy "Mit gondolnak az emberek az új Brad Pitt-filmről?"), és a legtöbbször Siri teljesíti is a kérést. Siri egy kis nem produktív fecsegést is elszórakoztat. Ha megkérdezed tőle, hogy mi az élet értelme, a "42"-et fogja válaszolni, amire *A Galaxis útikalauz stopposoknak* rajongói felismerhetnek, mint "az élet, az univerzum és minden végső kérdésére adott választ". A tudás kérdéseire (beleértve az élet értelmére vonatkozót is) a Wolfram Alpha válaszol, amelyet az [oldalón170](#) ismertetünk. A "chatbotok" egész világa létezik, akik nem csinálnak mást, csak csevegnek. Ha szeretne beszélgetni a Ramona nevű chatbotunkkal, látogasson el a KurzweilAI.net weboldalunkra, és kattintson a "Chat with Ramona" gombra.

Néhányan panaszkodtak nekem arra, hogy Siri nem válaszol bizonyos kérésekre, de gyakran eszembe jut, hogy ezek ugyanazok az emberek, akik kitartóan panaszkodnak az emberi szolgáltatókra is. Néha javaslom, hogy próbáljuk ki együtt, és gyakran jobban működik, mint amire számítottak. A panaszok a sakkozó kutya történetére emlékeztetnek. Egy hitetlenkedő kérdezőnek a kutya gazdája azt válaszolja: "Igen, igaz, sakkozik, de a végjátéka gyenge". Mostanában hatékony versenytársak jelennek meg, mint például a Google Voice Search.

Új korszakot jelent, hogy a nagyközönség ma már természetes beszélt nyelven beszélget a kézi számítógépeivel. Jellemző, hogy az emberek egy első generációs technológia jelentőségét annak korlátai miatt elvetik. Néhány évvel később, amikor a technológia már jól működik, az emberek még mindig elutasítják a jelentőségét, mert, nos, már nem új. Ettől függetlenül a Siri lenyűgözően működik egy első generációs termékhez képest, és egyértelmű, hogy ez a termékkategória csak egyre jobb lesz.

A Siri a Nuance HMM-alapú beszédfelismerő technológiáit használja. A természetes nyelvi kiterjesztéseket először a DARPA által finanszírozott "CALO" projekt fejlesztette ki.¹⁵ A Siri a Nuance saját természetes nyelvi technológiáival bővült, és a Nuance egy nagyon hasonló technológiát kínál Dragon Go!¹⁶ A természetes nyelv megértésére használt módszerek nagyon hasonlítanak a hierarchikus rejtett Markov-modellekre, sőt magát a HHMM-et is gyakran használják. Bár e rendszerek némelyikén nem szerepel kifejezetten az a felirat, hogy HMM-et vagy HHMM-et használ, a matematika gyakorlatilag azonos. Mindegyik lineáris szekvenciák hierarchiáját foglalja magában, ahol minden egyes elemnek súlya van, az összeköttetések önbefolyásolóak, és egy olyan általános rendszer, amely a tanulási adatok alapján önszerveződik. A tanulás általában a rendszer tényleges használata

során folytatódik. Ez a megközelítés megfelel a természetes nyelv hierarchikus szerkezetének - ez csak egy természetes folytatás a fogalmi létrán felfelé a beszédrészekről a szavakon át a mondatokon át a szemantikai struktúrákig. A hierarchikus tanulási rendszerek ezen osztályának pontos tanulási algoritmusát vezérlő paramétereken érdemes lenne genetikus algoritmust futtatni, és meghatározni az optimális algoritmikus részleteket.

Az elmúlt évtizedben változott a hierarchikus struktúrák kialakításának módja. Douglas Lenat (1950-ben született) 1984-ben indította el a nagyszabású Cyc (enCYClopedic) projektet, amelynek célja olyan szabályok létrehozása volt, amelyek kodifikálják a mindennapi "józan ész" ismereteit. A szabályok hatalmas hierarchiába szerveződtek, és minden egyes szabály - ismét - állapotok lineáris sorozatát tartalmazta. Például egy Cyc-szabály kimondhatja, hogy a kutyának arca van. A Cyc aztán összekapcsolhatja az arcok szerkezetére vonatkozó általános szabályokat: hogy egy arcnak két szeme, egy orra és egy szája van, és így tovább. Nem szükséges, hogy egy szabályrendszer legyen a kutya arcára, majd egy másik a macska arcára, bár természetesen további szabályokat is beiktathatunk arra vonatkozóan, hogy a kutyák arca miben különbözik a macskákétól. A rendszer tartalmaz egy következtető motort is: ha vannak olyan szabályaink, amelyek szerint a cocker spániel egy kutya, a kutyák állatok, és az állatok esznek ételt, és ha megkérdeznénk a következtető motort, hogy a cocker spánielek esznek-e, a rendszer azt válaszolná, hogy igen, a cocker spánielek esznek ételt. A következő hús év alatt, több ezer emberévnyi erőfeszítéssel több mint egymillió ilyen szabályt írtunk és teszteltünk. Érdekes módon a Cyc szabályok írására szolgáló nyelv - a CycL - szinte teljesen megegyezik a LISP nyelvvel.

Eközben egy ellenkező irányzat úgy vélte, hogy a természetes nyelvi megértés és általában az intelligens rendszerek létrehozásának legjobb megközelítése a rendszer által elsajátítani kívánt jelenségek nagyon nagy számú példányának kitettségből származó automatikus tanulás. Az ilyen rendszer egyik leghatásosabb példája a Google Fordító, amely ötven nyelvre és nyelvről tud fordítani. Ez 2500 különböző fordítási irányt jelent, bár a legtöbb nyelvpár esetében ahelyett, hogy a nyelvet közvetlenül a 2. nyelvre fordítaná, az 1. nyelvet angolra fordítja, majd az angolt a nyelvre Ez kilencvennyolcra 2.csökkenti a Google fordítóprogramjainak számát (plusz korlátozott számú nem angol nyelvpár, amelyek esetében van közvetlen fordítás). A Google fordítói nem használnak nyelvtani szabályokat, hanem a két nyelv között lefordított dokumentumok nagy "Rosetta-kő" korpuszai alapján hatalmas adatbázisokat hoznak létre minden egyes nyelvpárhoz a közös fordításokból. Az ENSZ hivatalos nyelveit alkotó hat nyelv esetében a Google az ENSZ dokumentumait használta, mivel azokat mind a hat nyelven közzéteszik. A kevésbé elterjedt nyelvek esetében más forrásokat használtak.

Az eredmények gyakran lenyűgözőek. A DARPA évente versenyeket rendez a különböző nyelvpárok legjobb automatizált nyelvi fordítórendszereinek kiválasztására, és a Google Translate gyakran győz bizonyos nyelvpárok esetében, megelőzve a közvetlenül emberi nyelvészek által létrehozott rendszereket.

Az elmúlt évtizedben két fontos felismerés mélyen befolyásolta a természetes nyelvi megértés területét. Az első a hierarchiákkal kapcsolatos. Bár a Google-megközelítés az

egyik nyelvből a másikba átvitt sima szószorozatok asszociációjával indult, a nyelv eredendően hierarchikus jellege elkerülhetetlenül belopta magát a működésébe. A hierarchikus tanulást módszeresen beépítő rendszerek (például hierarchikus rejtett Markov-modellek) lényegesen jobb teljesítményt nyújtottak. Az ilyen rendszereket azonban nem olyan automatikus felépíteni. Ahogyan az embereknek egyszerre körülbelül egy fogalmi hierarchiát kell megtanulniuk, ugyanez igaz a számítógépes rendszerekre is, így a tanulási folyamatot gondosan kell irányítani.

A másik felismerés az, hogy a kézzel alkotott szabályok jól működnek a közös alapismeretek magja esetében. A rövid szövegek fordításánál ez a megközelítés gyakran pontosabb eredményeket ad. A DARPA például a szabályalapú kínai-angol fordítókat rövid szövegek esetében magasabbra értékelte, mint a Google Translate-et. Az úgynevezett "nyelv farka" esetében, amely a nyelvben használt ritkán előforduló kifejezések és fogalmak millióira utal, a szabályalapú rendszerek pontossága elfogadhatatlanul alacsony aszimptotumhoz közelít. Ha a természetes nyelvek megértésének pontosságát az elemzett képzési adatok mennyiségével vetítjük össze, a szabályalapú rendszerek kezdetben nagyobb teljesítményt nyújtanak, de a meglehetősen alacsony, 70 százalék körüli pontossági szintre süllyednek. Ezzel éles ellentétben a statisztikai rendszerek elérhetik a magas 90-es pontosságot, de ehhez nagy mennyiségű adatra van szükség.

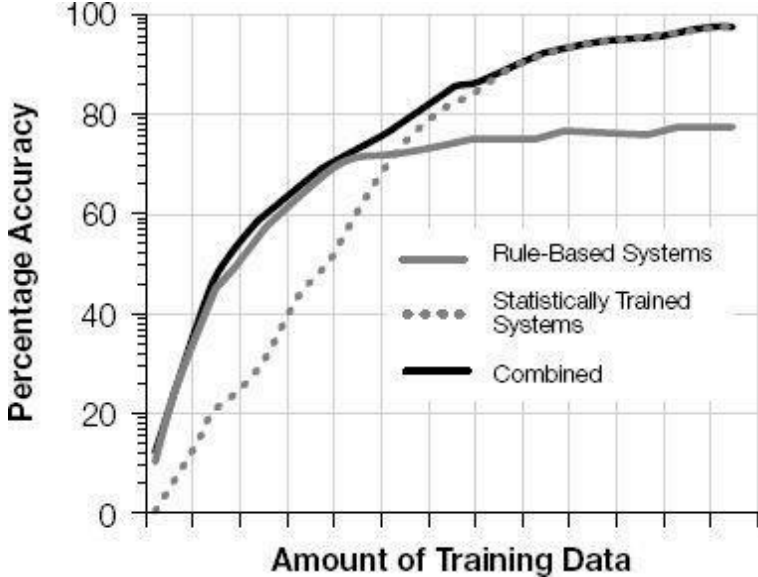
Gyakran szükségünk van a kis mennyiségű képzési adaton elért legalább mérsékelt teljesítmény kombinációjára, majd a lehetőségre, hogy nagyobb mennyiséggel nagy pontosságot érjünk el. A mérsékelt teljesítmény gyors elérése lehetővé teszi, hogy a rendszert a terepre helyezzük, majd automatikusan gyűjtsünk képzési adatokat, ahogy az emberek ténylegesen használják. Ily módon a rendszer használatával egyidejűleg nagymértékű tanulás történhet, és a pontossága javulni fog. A statisztikai tanulásnak teljesen hierarchikusnak kell lennie, hogy tükrözze a nyelv természetét, ami az emberi agy működését is tükrözi.

Így működik a Siri és a Dragon Go! is - a leggyakoribb és legmegbízhatóbb jelenségekre vonatkozó szabályokat használva, majd a valódi felhasználók kezében megtanulják a nyelv "farkát". Amikor a Cyc csapata rájött, hogy a kézzel kódolt szabályokon alapuló teljesítményük elérte a plafont, ők is ezt a megközelítést alkalmazták. A kézzel kódolt szabályok két alapvető funkciót látnak el. Megfelelő kezdeti pontosságot biztosítanak, hogy a próbarendszer széleskörű használatba kerülhessen, ahol automatikusan javulni fog. Másodszor, szilárd alapot biztosítanak a fogalmi hierarchia alacsonyabb szintjeihez, hogy az automatikus tanulás megkezdhesse a magasabb fogalmi szintek megtanulását.

Amint fentebb említettük, Watson különösen látványos példája a kézzel kódolt szabályok és a hierarchikus statisztikai tanulás kombinálásának. Az IBM számos vezető természetes nyelvi programot kombinált, hogy létrehozson egy olyan rendszert, amely képes volt a *Jeopardy!* 2011. február 14-16-án Watson megmérkőzött a két vezető emberi játékosal: Brad Rutterrel, aki mindenkinél több pénzt nyert a kvízműsorban, és Ken Jenningszel, aki korábban hetvenöt napig tartotta a *Jeopardy!* bajnoki címét.

A kontextus kedvéért: az 1980-as évek közepén írt első könyvemben, *Az intelligens*

gépek korában azt jósoltam, hogy 1998-ra egy számítógép fogja megnyerni a sakkvilágbajnokságot. Azt is megjósoltam, hogy amikor ez megtörténik, vagy leértékeljük az emberi intelligenciáról alkotott véleményünket, vagy felértékeljük a gépi intelligenciáról alkotott véleményünket, vagy leértékeljük a sakk jelentőségét, és ha a történelem útmutatásul szolgál, akkor a sakkot lekicsinyeljük. Mindkét dolog megtörtént 1997-ben. Amikor az IBM sakkszuperszámítógépe, a Deep Blue legyőzte az uralkodó emberi sakkvilágbajnokot, Garri Kaszparovot, azonnal olyan érveket hallhattunk, hogy várható volt, hogy egy számítógép győzni fog a sakkban, mert a számítógépek logikai gépek, és a sakk végül is egy logikai játék. Így a Deep Blue győzelmét nem ítélték sem meglepőnek, sem jelentősnek. Sok kritikusa azzal érvelt, hogy a számítógépek soha nem fogják elsajátítani az emberi nyelv finomságait, beleértve a metaforákat, hasonlatokat, szójátékokat, kétértelműségeket és a humort.



A természetes nyelvet megértő rendszerek pontossága a képzési adatok mennyiségének függvényében. A legjobb megközelítés a nyelv "magjára" vonatkozó szabályok és a nyelv "farkára" vonatkozó adatalapú megközelítés kombinálása.

Legalábbis ez az egyik oka annak, hogy Watson ilyen jelentős mérföldkőnek számít:

A Jeopardy! pontosan ilyen kifinomult és kihívást jelentő nyelvi feladat. *A Jeopardy!* tipikus kérdései az emberi nyelv számos ilyen szeszélyét tartalmazzák. Ami talán sok megfigyelő számára nem nyilvánvaló, az az, hogy Watsonnak nemcsak a váratlan és bonyolult lekérdezések nyelvét kellett elsajátítania, hanem a tudását nagyrészt nem kézzel kódolták. Ezt a tudást úgy szerezte meg, hogy 200 millió oldalnyi természetes nyelvű dokumentumot olvasott el, beleértve a Wikipédia és más enciklopédiák egészét is, ami trillió4 bájtnyi nyelvi alapú tudást tartalmaz. Amint azt e könyv olvasói jól tudják, a Wikipédia nem LISP-ben vagy CycL-ben íródott, hanem természetes mondatokban, amelyek a nyelvben rejlő összes kétértelműséggel és bonyolultsággal rendelkeznek. Watsonnak egy kérdés megválaszolásakor figyelembe kellett vennie mind a 4 billió karaktert a referenciaanyagában. (Tisztában vagyok vele, hogy a *Jeopardy!* lekérdezések kérdésre adott válaszok, de ez csak technikai kérdés - végső soron ezek valóban kérdések.) Ha Watson képes megérteni és megválaszolni a 200 millió oldal alapján feltett kérdéseket - három másodperc alatt! -, akkor semmi sem akadályozza meg, hogy a hasonló rendszerek elolvassák a világhálón található többi milliárd dokumentumot. Ez a törekvés már folyamatban van.

Amikor az 1970-es évektől az 1990-es évekig karakter- és beszédfelismerő rendszereket, valamint korai természetesnyelv-megértő rendszereket fejlesztettünk, a "szakértő menedzser" bevonásának módszerét alkalmaztuk. Több rendszert fejlesztettünk ki ugyanarra a feladatra, de mindegyikbe némileg eltérő megközelítéseket építettünk be. A különbségek némelyike finom volt, például a tanulási algoritmus matematikáját vezérlő

paraméterek eltérései. Egyes eltérések alapvetőek voltak, mint például a hierarchikus statisztikai tanulási rendszerek helyett a szabályalapú rendszerek bevonása. A szakértői menedzser maga is egy szoftverprogram volt, amelyet úgy programoztak, hogy valós helyzetekben vizsgálva teljesítményüket, megismerje e különböző rendszerek erősségeit és gyengeségeit. A program azon az elképzelésen alapult, hogy ezek az erősségek ortogonálisak, azaz az egyik rendszer ott lesz erős, ahol a másik gyenge. Valóban, a kiképzett szakértő menedzserrel a kezében a kombinált rendszerek összteljesítménye sokkal jobb volt, mint bármelyik egyedi rendszeré.

Watson ugyanígy működik. Az UIMA (Unstructured Information Management Architecture) elnevezésű architektúrát használva Watson szó szerint több száz különböző rendszert alkalmaz - a Watsonban található egyes nyelvi komponensek közül sok ugyanaz, mint amit a nyilvánosan elérhető természetesnyelv-értelmező rendszerek is használnak -, amelyek mindegyike vagy közvetlenül próbál választ adni a *Jeopardy!* lekérdezésére, vagy legalábbis megpróbálja a lekérdezést valamilyen módon megfejteni. Az UIMA alapvetően szakértői menedzserként működik, hogy a független rendszerek eredményeit intelligensen kombinálja. Az UIMA lényegesen túlmutat a korábbi rendszereken, például azon, amelyet a Nuance előd vállalatánál fejlesztettünk ki, mivel az egyes rendszerek hozzájárulhatnak az eredményhez anélkül, hogy feltétlenül végleges választ adnának. Elég, ha egy alrendszer segít leszűkíteni a megoldást. Az UIMA azt is ki tudja számítani, hogy mekkora a bizalma a végső válaszban. Az emberi agy is ezt teszi

-valószínűleg nagyon magabiztosan válaszolunk, ha anyánk keresztnevét kérdezik tőlünk, de kevésbé vagyunk azok, akiknek a nevét egy évvel ezelőtt véletlenül találkoztunk.

Az IBM tudósai tehát ahelyett, hogy egyetlen elegáns megközelítést találtak volna ki a *Jeopardy!* nyelvi problémájának megértésére, az IBM tudósai kombinálták az összes legkorszerűbb nyelvmegértő modult, amelyhez hozzájutottak. Egyesek hierarchikus rejtett Markov-modelleket használnak; mások a HHMM matematikai változatait; mások szabályalapú megközelítéseket alkalmaznak, hogy közvetlenül kódoljanak egy megbízható szabályokból álló alapkészletet. Az UIMA értékeli az egyes rendszerek teljesítményét a tényleges használat során, és optimális módon kombinálja őket. A Watsonról szóló nyilvános vitákban van némi félreértés abban a tekintetben, hogy az IBM tudósai, akik létrehozták, gyakran az UIMA-ra összpontosítanak, amely az általuk létrehozott szakértői menedzser. Ez egyes megfigyelők részéről olyan megjegyzésekhez vezet, hogy Watson nem érti igazán a nyelvet, mert nehéz azonosítani, hogy hol van ez a megértés. Bár az UIMA keretrendszer is tanul a saját tapasztalataiból, Watson nyelvi "megértése" nem csak az UIMA-ban található meg, hanem inkább annak számos összetevője között oszlik meg, beleértve az önszerveződő nyelvi modulokat, amelyek a HHMM-hez hasonló módszereket használnak.

Watson különálló része a technológia az UIMA által a válaszokra adott megbízhatósági becslést használja a *Jeopardy!* tétek elhelyezésének meghatározásához. Bár a Watson rendszer kifejezetten erre a bizonyos játékra van optimalizálva, alapvető nyelvi és tudáskeresési technológiája könnyen adaptálható más, széleskörű feladatokra. Azt gondolhatnánk, hogy a kevésbé elterjedt szakmai ismereteket, mint például az orvosi területen, nehezebb elsajátítani, mint az általános célú "általános" ismereteket, amelyek a *Jeopardy!* Valójában éppen az ellenkezője a helyzet: A szakmai tudás általában jobban

szervezett, strukturáltabb és kevésbé kétértelmű, mint a köznapit tudás, így nagyon jól alkalmazható a pontos természetes nyelvi megértésre az említett technikák segítségével. Mint említettük, az IBM jelenleg a Nuance-szal együttműködve dolgozik azon, hogy a Watson technológiát az orvostudományra adaptálja.

A beszélgetés, amely akkor zajlik, amikor Watson a *Jeopardy!* játékot játssza, rövid: Felteszünk egy kérdést, és Watson előáll egy válasszal. (Gyakorlatilag megint csak egy kérdéssel válaszol egy válaszra.) Nem vesz részt olyan beszélgetésben, amely az összes résztvevő összes korábbi kijelentésének nyomon követését igényelné. (A Siri valójában ezt korlátozott mértékben meg is teszi: Ha megkéri, hogy küldjön üzenetet a feleségének, megkéri, hogy azonosítsa őt, de a későbbi kéréseknél megjegyzi, hogy ki ő). Egy beszélgetés során az összes információ nyomon követése - ez a feladat egyértelműen szükséges lenne a Turing-teszt teljesítéséhez - jelentős többletkövetelmény, de alapvetően nem nehezebb, mint amit Watson már most is csinál. Végül is Watson több százmillió oldalnyi anyagot olvasott el, ami nyilvánvalóan sok történetet tartalmaz, így képes bonyolult, egymásra épülő események nyomon követésére. Ezért képesnek kell lennie arra, hogy nyomon kövesse a saját beszélgetéseit, és ezt figyelembe vegye a későbbi válaszaiban.

A *Jeopardy!* játék másik korlátja, hogy a válaszok általában rövidek: Például nem tesz fel olyan kérdéseket, mint amilyenek az *A Tale of Two Cities* öt fő témáját kéri a versenyzőktől. Amennyiben képes olyan dokumentumokat találni, amelyek a regény témáit tárgyalják, a Watson megfelelően módosított változatának képesnek kell lennie arra, hogy erre válaszoljon. Az már más kérdés, hogy önmagában, pusztán a könyv elolvasása alapján jöjjön rá ezekre a témákra, és ne lényegében más gondolkodók gondolatait másolja (akár szavak nélkül is). Ez magasabb szintű feladatot jelentene, mint amire Watson ma képes - ezt nevezem én Turing-teszt szintű feladatnak. (Ezzel együtt megjegyzem, hogy a legtöbb ember sem saját eredeti gondolatokkal áll elő, hanem lemásolja társai és véleményformálói gondolatait.) Mindenesetre 2012-ben járunk, nem 2029-ben, így még nem várnék Turing-teszt-szintű intelligenciát. Másfelől viszont felhívnám a figyelmet arra, hogy az olyan kérdésekre adott válaszok értékelése, mint például egy regény kulcsgondolatainak megtalálása, önmagában nem egyszerű feladat. Ha valakitől megkérdezik, hogy ki írta alá a Függetlenségi Nyilatkozatot, meg lehet állapítani, hogy a válasza igaz vagy hamis. Az olyan magasabb szintű kérdésekre adott válaszok érvényességét, mint például egy kreatív mű témáinak leírása, sokkal nehezebb megállapítani.

Figyelemre méltó, hogy bár Watson nyelvtudása valójában valamivel elmarad egy képzett emberétől, mégis képes volt legyőzni a világ két legjobb *Jeopardy!* játékosát. Ezt azért tudta elérni, mert képes kombinálni nyelvi képességeit és tudásmegértését a gépek által birtokolt tökéletes emlékezőképességgel és rendkívül pontos memóriával. Ezért is van az, hogy személyes, társadalmi és történelmi emlékeinket már nagyrészt rájuk bíztuk.

Bár nem vagyok hajlandó feljebb vinni a Turing-tesztet átmenő számítógépre vonatkozó előrejelzésemet, a Watsonhoz hasonló rendszerekben elért 2029.fejlődésnek bárki számára komoly bizalmat kellene adnia, hogy a Turing-szintű mesterséges intelligencia eljövetele közel van. Ha valaki létrehozná a Watson egy olyan változatát, amelyet a Turing-tesztre optimalizáltak, valószínűleg elég közel kerülne hozzá.

John Searle amerikai filozófus (született 1932-ben) nemrég azt állította, hogy Watson

nem képes gondolkodni. A "kínai szoba" című gondolatkísérletére hivatkozva (amelyet a [11. fejezetben](#) fogok bővebben tárgyalni) azt állítja, hogy Watson csak szimbólumokkal manipulál, és nem érti a szimbólumok jelentését. Valójában Searle nem írja le pontosan Watsont, mivel a nyelv megértése hierarchikus statisztikai folyamatokon alapul - nem pedig szimbólumok manipulálásán. Searle jellemzése csak akkor lenne pontos, ha Watson önszerveződő folyamatainak minden lépését "szimbólumok manipulálásának" tekintenénk. De ha ez lenne a helyzet, akkor az emberi agyat sem ítélnénk gondolkodásra képesnek.

Mulatságos és ironikus, amikor a megfigyelők kritizálják Watsont, amiért *csak* statisztikai elemzéseket végez a nyelvről, szemben azzal, hogy az emberek "igazi" nyelvi megértése birtokában van. A hierarchikus statisztikai elemzés pontosan az, amit az emberi agy csinál, amikor több hipotézist old meg statisztikai következtetés alapján (és valóban a neokortikális hierarchia minden szintjén). Watson és az emberi agy egyaránt a hierarchikus megértés hasonló megközelítése alapján tanul és reagál. Watson tudása sok tekintetben sokkal kiterjedtebb, mint az emberé; egyetlen ember sem állíthatja, hogy a Wikipédia egészét elsajátította, ami Watson tudásbázisának csak egy része. Ezzel szemben egy ember ma már több fogalmi szintet képes elsajátítani, mint Watson, de ez bizonyosan nem jelent állandó szakadékot.

Az egyik fontos rendszer, amely a szervezett tudásra alkalmazott számítástechnika erejét mutatja, a Wolfram Alpha, a Wolfram Research brit matematikus és tudós, Dr. Wolfram (szül. 1959) és munkatársai által kifejlesztett válaszmotor (szemben a keresőmotorral). Ha például megkérdezzük a Wolfram Alfát (a WolframAlpha.com oldalon), hogy "Hány prímszám van egymillió alatt?", a válasz "78 498" lesz. A választ nem kikereste, hanem kiszámította, és a válasz után közli az általa használt egyenleteket. Ha ezt a választ egy hagyományos keresőmotor segítségével próbálnád megszerezni, akkor olyan linkekre irányítana, ahol megtalálhatnád a szükséges algoritmusokat. Ezeket a képleteket aztán be kellene illesztened egy olyan rendszerbe, mint a Mathematica, amelyet szintén Dr. Wolfram fejlesztett ki, de ez nyilvánvalóan sokkal több munkát (és megértést) igényel, mintha egyszerűen csak megkérdeznéd az Alfát.

Az Alpha valóban több millió15 sor Mathematica kódból áll. Az Alpha szó szerint kiszámítja a választ körülbelül 10 trillió bájtnyi adatból, amelyet a Wolfram Research munkatársai gondosan kurátori kezelésbe vettek. Tényszerű kérdések széles skáláját teheti fel, például "Melyik országnak van a legmagasabb az egy főre jutó GDP-je?". (Válasz: Monaco, ahol egy főre 212 000 dollár jut amerikai dollárban), vagy "Hány éves Stephen Wolfram?". (Válasz: 52 év, 9 hónap, nap2, amikor ezt írom). Mint említettük, az Alfát az Apple Siri részeként használják; ha tényszerű kérdést teszünk fel a Sirinek, azt átadja az Alfának, hogy kezelje. Az Alpha kezeli a Microsoft Bing keresőmotorjának feltett keresések egy részét is.

Dr. Wolfram egy nemrégiben megjelent blogbejegyzésében arról számolt be, hogy az Alpha az esetek 90 százalékában sikeres válaszokat ad.¹⁷ Arról is beszámol, hogy a hibaarány exponenciálisan csökken, a felezési idő körülbelül másfél év. Ez egy lenyűgöző rendszer, és a kézzel készített módszerek és kézzel ellenőrzött adatok. Ez a bizonyíték arra, hogy miért is

hoztuk létre a számítógépeket. Ahogyan felfedezzük és összeállítjuk a tudományos és matematikai módszereket, a számítógépek sokkal jobbak a segítség nélküli emberi intelligenciánál a megvalósításukban. Az ismert tudományos módszerek nagy részét az Alfába kódolták, a folyamatosan frissülő adatokkal együtt a közgazdaságtantól a fizikáig terjedő témákban. Egy Dr. Wolframmal folytatott magánbeszélgetésem során úgy becsülte, hogy az önszerveződő módszerek, mint amilyeneket a Watsonban használnak, jellemzően 80 százalékos pontosságot érnek el, ha jól működnek. Az Alpha - mutatott rá - körülbelül 90 százalékos pontosságot ér el. Természetesen mindkét pontossági számban van önszelekción, mivel a felhasználók (mint például én magam) megtanulták, hogy az Alpha milyen típusú kérdésekben jó, és hasonló tényező vonatkozik az önszerveződő módszerekre is. A nyolcvan százalékos becslésnek tűnik arra vonatkozóan, hogy a Watson mennyire pontos a *Jeopardy!* kérdésekben, de ez elegendő volt ahhoz, hogy legyőzze a legjobb embereket.

Véleményem szerint az olyan önszerveződő módszerekre, mint amilyeneket az elme mintafelismerési elméletében megfogalmaztam, szükség van a bonyolult és gyakran kétértelmű hierarchiák megértéséhez, amelyekkel a valós világ jelenségeiben, köztük az emberi nyelvben találkozunk. Egy robusztusan intelligens rendszer ideális kombinációja az lenne, ha a PRTM-en alapuló hierarchikus intelligenciát (amely szerintem az emberi agy működése) a tudományos ismeretek és adatok pontos kodifikációjával kombinálnánk. Ez lényegében egy embert ír le egy számítógéppel. Az elkövetkező években az intelligencia mindkét pólusát fejleszteni fogjuk. Ami a biológiai intelligenciánkat illeti, bár neokortexünk jelentős plaszticitással rendelkezik, alapvető felépítését fizikai korlátai korlátozzák. Az, hogy további neokortexet helyeztünk a homlokunkba, fontos evolúciós innováció volt, de most nem tudjuk könnyen ezerszeresére, vagy akár 10 százalékkal is növelni a homloklebenyünk méretét. Vagyis biológiailag nem tudjuk megtenni, de technológiailag pontosan ezt fogjuk tenni.

Stratégia az elme megteremtéséhez

Az agyunkban több milliárd neuron található, de mik is azok a neuronok? Csak sejtek. Az agynak nincs tudása, amíg a neuronok között nem jönnek létre kapcsolatok. Minden, amit tudunk, minden, ami vagyunk, abból származik, ahogyan a neuronjaink összekapcsolódnak.

-Tim Berners-Lee

Használjuk fel a fentebb tárgyalt megfigyeléseimet, hogy elkezdjük az agy felépítését. Kezdjük egy olyan mintafelismerő építésével, amely megfelel a szükséges tulajdonságoknak. Ezután annyi másolatot készítünk a felismerőből, ahány memóriával és számítási erőforrással rendelkezünk. Minden felismerő kiszámítja annak a valószínűségét,

hogy az adott mintát felismerték. Ennek során figyelembe veszi az egyes bemenetek megfigyelt nagyságát (valamilyen megfelelő kontinuumban), és ezeket összeveti az egyes bemenetekhez tartozó tanult méret- és méretváltózási paraméterekkel. A felismerő akkor indítja el a szimulált axonját, ha ez a kiszámított valószínűség meghalad egy küszöbértéket. Ez a küszöbérték és a minta valószínűségének kiszámítását vezérlő paraméterek azok közé a paraméterek közé tartoznak, amelyeket genetikai algoritmussal optimalizálunk. Mivel a minta felismeréséhez nem szükséges, hogy minden bemenet aktív legyen, ez lehetővé teszi az autoasszociatív felismerést (azaz a minta felismerését a minta csak egy részének jelenléte alapján). Lehetővé tesszük a gátló jeleket is (olyan jelek, amelyek azt jelzik, hogy a minta kevésbé valószínű).

A minta felismerése aktív jelet küld felfelé a mintafelismerő szimulált axonján. Ez az axon viszont egy vagy több mintafelismerőhöz kapcsolódik a következő, magasabb fogalmi szinten. A következő magasabb fogalmi szinten csatlakozó összes mintafelismerő elfogadja ezt a mintát az egyik bemenetként.

Minden mintafelismerő jeleket küld lefelé az alacsonyabb fogalmi szinteken lévő mintafelismerőknek, amikor egy minta nagy részét felismerte, jelezve, hogy a minta többi része "várható". Minden mintafelismerő rendelkezik egy vagy több ilyen várható jel bemeneti csatornával. Amikor egy ilyen módon várt jel érkezik, akkor az adott mintafelismerő felismerési küszöbértéke csökken (könnyebbé válik).

A mintafelismerők felelősek azért, hogy "összekapcsolódjanak" más mintafelismerőkkel a fogalmi hierarchiában felfelé és lefelé. Megjegyzendő, hogy a szoftveres megvalósításban az összes "vezeték" virtuális kapcsolatokon keresztül működik (amelyek a webes kapcsolatokhoz hasonlóan alapvetően memóriamutatók), nem pedig tényleges vezetékeken keresztül. Ez a rendszer valójában sokkal rugalmasabb, mint a biológiai agyé. Az emberi agyban az új mintákat egy tényleges fizikai mintafelismerőhöz kell rendelni, és az új kapcsolatokat egy tényleges axon-dendrit kapcsolaton keresztül kell létrehozni. Ez általában azt jelenti, hogy egy meglévő fizikai kapcsolatot veszünk, amely megközelítőleg megfelel a szükségesnek, majd megnöveljük a szükséges axon- és dendritnyúlványokat, hogy a teljes kapcsolat létrejöhessen.

Egy másik, biológiai emlősök agyában alkalmazott technika az, hogy a lehetséges kapcsolatok nagy számából indulunk ki, majd a nem használt neurális kapcsolatokat megkurtítjuk. Ha egy biológiai neokortex átcsoportosítja a már régebbi mintákat megtanult agykérgi mintafelismerőket, hogy újabb anyagot tanuljanak meg, akkor a kapcsolatokat fizikailag át kell konfigurálni. Ismétlem, ezek a feladatok sokkal egyszerűbbek egy szoftveres megvalósításban. Egyszerűen új memóriahelyeket rendelünk egy új mintafelismerőhöz, és memóriakapcsolatokat használunk a kapcsolatokhoz. Ha a digitális neokortex át akarja rendelni az agykérgi memóriaforrásokat az egyik mintakészletről a másikra, akkor egyszerűen visszahelyezi a régi mintafelismerőket a memóriába, majd elvégzi az új hozzárendelést. Ez a fajta "szemétyűjtés" és a memória újbóli hozzárendelése számos szoftverrendszer architektúrájának szabványos jellemzője. A digitális agyunkban a régi emlékeket is lementenénk, mielőtt kidobnánk őket az aktív neokortexből, ezt az óvintézkedést biológiai agyunkban nem tudjuk megtenni.

Az önszerveződő hierarchikus mintafelismerés ezen megközelítésének megvalósításához számos matematikai technika alkalmazható. Az általam használt módszer a hierarchikus rejtett Markov-modellek, több okból is. Személy szerint több évtizedes ismerettséggel rendelkezem ezzel a módszerrel, mivel az 1980-as évektől kezdve a legkorábbi beszédfelismerő és természetes nyelvi rendszerekben használtam. A teljes szakterület szempontjából a rejtett Markov-modellekkel nagyobb tapasztalat áll rendelkezésre, mint bármely más megközelítéssel a mintafelismerési feladatok esetében. A természetes nyelvi megértésben is széles körben használják őket. Számos NLU-rendszer használ olyan technikákat, amelyek legalább matematikailag hasonlóak a HHMM-hez.

Megjegyzendő, hogy nem minden rejtett Markov-modell rendszer teljesen hierarchikus. Egyesek csak néhány hierarchiaszintet tesznek lehetővé - például az akusztikai állapotoktól a fonémákon át a szavakig. Egy agy felépítéséhez lehetővé akarjuk tenni a rendszerünk számára, hogy annyi új hierarchiaszintet hozzon létre, amennyit csak szükséges. Emellett a legtöbb rejtett Markov-modell rendszer nem teljesen önszerveződő. Némelyiknek rögzített kapcsolatai vannak, bár ezek a rendszerek hatékonyan metszik sok kiindulási kapcsolatukat azáltal, hogy lehetővé teszik számukra, hogy nulla kapcsolati súlyokat fejlesszenek ki. Az 1980-as és 1990-es évekbeli rendszereink automatikusan metszették a bizonyos szint alatti kapcsolati súlyokkal rendelkező kapcsolatokat, és lehetővé tették új kapcsolatok létrehozását is a képzési adatok jobb modellezése és a menet közbeni tanulás érdekében. Úgy vélem, kulcsfontosságú követelmény, hogy a rendszer rugalmasan létrehozassa saját topológiáit a tanulás során tapasztalt minták alapján. A lineáris programozás matematikai technikáját használhatjuk a kapcsolatok optimális hozzárendelésére az új mintafelismerőkhöz.

Digitális agyunk az egyes minták jelentős redundanciáját is befogadja, különösen a gyakran előforduló mintákét. Ez lehetővé teszi a gyakori minták robusztus felismerését, és egyben az egyik legfontosabb módszer a minta különböző formáinak invariáns felismeréséhez. Szükségünk lesz azonban szabályokra arra vonatkozóan, hogy mennyi redundanciát engedjünk meg, mivel nem akarjuk, hogy a nagyon gyakori, alacsony szintű minták túlzott mennyiségű memóriát használjanak fel.

A redundanciára vonatkozó szabályok, a felismerési küszöbértékek és az "ez a minta várható" jelzés hatása a küszöbértékre néhány példa az ilyen típusú önszerveződő rendszer teljesítményét befolyásoló kulcsfontosságú általános paraméterekre. Ezeket a paramétereket kezdetben intuíción alapján állítanám be, de aztán genetikai algoritmus segítségével optimalizálnánk őket.

Nagyon fontos szempont az agy képzése, legyen az biológiai vagy szoftveres. Amint azt korábban már említettem, egy hierarchikus mintafelismerő rendszer (digitális vagy biológiai) egyszerre csak két - lehetőleg egy - hierarchikus szintet fog megtanulni. A rendszer feltöltéséhez olyan, korábban betanított hierarchikus hálózatokkal kezdeném, amelyek már megtanulták a leckét az emberi beszéd, a nyomtatott betűk és a természetes nyelvi struktúrák felismerésében. Egy ilyen rendszer képes lenne természetes nyelvű dokumentumok olvasására, de egyszerre csak megközelítőleg egy fogalmi szintet tudna elsajátítani. A korábban megtanult szintek viszonylag stabil alapot biztosítanak a következő szint megtanulásához. A rendszer újra és újra el tudja olvasni ugyanazokat a

dokumentumokat, és minden egyes olvasáskor új fogalmi szintekre tesz szert, hasonlóan ahhoz, ahogyan az emberek újraolvassák és mélyebb megértést érnek el a szövegekből. A világhálón több milliárd oldalnyi anyag érhető el. Maga a Wikipédia angol nyelvű változata körülbelül négy millió cikket tartalmaz. Egy kritikus gondolkodást segítő modult is biztosítanék, amely folyamatosan háttérben vizsgálná az összes létező mintát, és felülvizsgálná azok kompatibilitását a szoftver neokortexében lévő többi mintával (ötlettel). A biológiai agyunkban nincs ilyen lehetőség, ezért van az, hogy az emberek képesek teljesen ellentmondásos gondolatokat is nyugodtan tartani. Egy ellentmondásos gondolat azonosítása után a digitális modul elkezdene a megoldás keresését, beleértve a saját agykérgi struktúráit, valamint a rendelkezésére álló hatalmas irodalmat. A feloldás jelentheti egyszerűen annak megállapítását, hogy az egyik ellentmondásos gondolat egyszerűen helytelen (ha az ellentmondó adatok túlnyomó része ezt ellenzi). Konstruktívabb lenne, ha egy magasabb fogalmi szinten találna egy olyan elképzelést, amely feloldja a látszólagos ellentmondást azáltal, hogy olyan perspektívát nyújt, amely mindkét elképzelést megmagyarázza. A rendszer ezt a feloldást új mintaként hozzáadná, és összekapcsolná a feloldás keresését eredetileg kiváltó gondolatokkal. Ez a kritikai gondolkodási modul folyamatos háttérfeladatként futna. Nagyon hasznos lenne, ha az emberi agyak ugyanezt tennék.

Olyan modult is biztosítanék, amely minden tudományágban azonosítja a nyitott kérdéseket. További folyamatos háttérfeladatként ezekre megoldást keresne a tudás más, egymástól eltérő területein. Mint említettem, a neokortexben lévő tudás mélyen egymásba ágyazott mintázatokból áll, és ezért teljes mértékben metaforikus. Egy-egy mintát felhasználhatunk arra, hogy megoldást vagy felismerést nyújtson egy látszólag összefüggéstelen területen.

Példaként emlékezzünk vissza a [4. fejezet](#)ben használt metaforára, amely a gázban lévő molekulák véletlenszerű mozgását az evolúciós változások véletlenszerű mozgásához kapcsolja. A molekulák a gázban véletlenszerűen mozognak, látszólagos irányérzék nélkül. Ennek ellenére gyakorlatilag minden gázmolekula egy főzőpohárban, ha elegendő időt kap, elhagyja a poharat. Megjegyeztem, hogy ez perspektívát nyújt az intelligencia evolúciójával kapcsolatos egyik fontos kérdésre. A gázban lévő molekulákhoz hasonlóan az evolúciós változások is mindenféleképpen mozognak, látszólagos irány nélkül. Ennek ellenére mégis látunk egy mozgást a nagyobb komplexitás és a nagyobb intelligencia felé, sőt az evolúció legfőbb vívmánya, a hierarchikus gondolkodásra képes neokortex kialakulása felé. Így egy másik terület (a termodinamika) vizsgálatával betekintést nyerhetünk abba, hogy egy látszólag céltalan és iránytalan folyamat hogyan érhet el egy látszólag céltudatos eredményt egy területen (a biológiai evolúció).

Korábban már említettem, hogy Charles Lyell felismerése, miszerint az áramló víz által a sziklaalakzatokban végbemenő apró változások idővel nagy völgyeket vájhatnak, Charles Darwint is hasonló megfigyelésre ösztönözte az egy fajon belüli organizmusok jellemzőinek folyamatos apró változásával kapcsolatban. Ez a metaforakeresés egy másik folyamatos háttérfolyamat lenne.

Egyszerre több listán is végig kell lépkednünk, hogy a strukturált gondolkodás megfelelőjét biztosítsuk. Egy lista lehet azoknak a feltételeknek a kijelentése, amelyeknek egy probléma megoldásának meg kell felelnie. Minden egyes lépés létrehozhat egy rekurzív

keresést a meglévő ötlethierarchiában vagy a rendelkezésre álló irodalomban való keresést. Úgy tűnik, hogy az emberi agy egyszerre csak négy egyidejű listát képes kezelni (olyan eszközök, mint a számítógépek segítségével), de nincs ok arra, hogy egy mesterséges neokortexnek ilyen korlátja legyen.

A mesterséges agyunkat is fel akarjuk majd vértetni a fajta intelligenciával, amelyben a számítógépek mindig is jeleskedtek, vagyis azzal a képességgel, hogy hatalmas adatbázisokat pontosan elsajátítsanak, és ismert algoritmusokat gyorsan és hatékonyan hajtsanak végre. A Wolfram Alpha egyedülálló módon számos ismert tudományos módszert egyesít, és alkalmazza azokat a gondosan összegyűjtött adatokra. Ez a fajta rendszer is tovább fog fejlődni, tekintettel Dr. Wolfram megfigyelésére, miszerint a hibaarányok exponenciálisan csökkennek.

Végül az új agyunknak szüksége van egy célra. A célt célok sorozataként fejezzük ki. Biológiai agyunk esetében a céljainkat az öröm- és félelemközpontok határozzák meg, amelyeket a régi agyunktól örököltünk. Ezeket a primitív késztetéseket eredetileg a biológiai evolúció állította be, hogy elősegítse a fajok túlélését, de a neokortex lehetővé tette számunkra, hogy szublimáljuk őket. Watson célja az volt, hogy válaszoljon a *Jeopardy!* kérdésekre. Egy másik, egyszerűen megfogalmazott cél az lehet, hogy átmenjen a Turingteszten. Ehhez a digitális agynak szüksége lenne egy emberi narratívára, a saját fiktív történetére, hogy biológiai embernek adhassa ki magát. Emellett jelentősen le kellene butítania önmagát, mert bármelyik rendszer, amelyik mondjuk Watson tudását mutatná, gyorsan lelepleződne, mint nem biológiai.

Még érdekesebb, hogy az új agyunknak egy ambiciózusabb célt is adhatnánk, például hozzájárulhatnánk egy jobb világhoz. Egy ilyen cél természetesen sok kérdést vet fel: Kinek jobb? Milyen módon jobb? A biológiai emberek számára? Minden tudatos lény számára? Ha ez a helyzet, akkor ki vagy mi a tudatos?

Ahogy a nem biológiai agyak ugyanolyan képessé válnak arra, hogy a biológiai agyakhoz hasonlóan változásokat idézzenek elő a világban - sőt, végső soron sokkal inkább képessé válnak arra, mint a biológiai agyak -, úgy kell gondolkodnunk az erkölcsi nevelésükről. Jó kiindulópont lehet a vallási hagyományaink egyik régi eszméje: az aranszabály.

FEJEZET 8

AZ ELME MINT SZÁMÍTÓGÉP

A francia parasztkenyérhez hasonló formájú agyunk egy zsúfolt kémiai laboratórium, ahol megállás nélkül zajlanak az idegi beszélgetések. Képzeljük el az agyat, ezt a fényes lényhalmot, ezt az egérszürke sejtparlamentet, ezt az álmogyárat, ezt a kis zsarnokot egy csontgolyó belsejében, ezt az idegsejtek halmazát, amelyik minden játéknak parancsol, ezt a kis mindenütt, ezt a szeszélyes örömkupolát, ezt az én gyűrött ruhatárát, amelyik úgy van a koponyába gyömöszölve, mint túl sok ruha a tornazsákba.

-Diane Ackerman

Az agyak azért léteznek, mert a túléléshez szükséges erőforrások eloszlása és a túlélést fenyegető veszélyek térben és időben változnak.

-John M. Allman

Az agy modern földrajza finoman elavultnak hat - inkább olyan, mint egy középkori térkép, ahol az ismert világot terra incognita veszi körül, ahol szörnyek kóborolnak.

-David Bainbridge

A matematikában nem érted a dolgokat. Csak megszokod őket.

-John von Neumann

A számítógépek huszadik század közepén történt megjelenése óta folyamatos vita folyik nemcsak a számítógépek képességeinek végső mértékéről, hanem arról is, hogy maga az emberi agy is tekinthető-e egyfajta számítógépnek. Ami az utóbbi kérdést illeti, a konszenzus a kétféle információfeldolgozó egység alapvetően azonosnak tekintésétől az alapvetően különbözőségükig terjedt. Az agy tehát számítógép?

Amikor a számítógépek az 1940-es években először váltak népszerűvé, azonnal gondolkodó gépeknek tekintették őket. Az 1946-ban bejelentett ENIAC-ot a sajtó "óriási agyként" jellemezte. Ahogy a következő évtizedben a számítógépek kereskedelmi forgalomban is elérhetővé váltak, a reklámok rendszeresen úgy hivatkoztak rájuk, mint agyakra, amelyek olyan teljesítményre képesek, amire a közönséges biológiai agyak nem képesek.

simple as a desk calculator...

intelligent as an electronic brain!

the **e101**

combine advantages of both... and solves the problems "man-made," at low cost and with big savings in technical man-hours.

Robust programming - an exclusive feature of the E101 digital computer - can be entered in its binary transfer rotators, so coding. That's why more E101's are in work than all other computers requires combined.

Immediate delivery, but not servicable. At a time, send us one of your problems. We'll program it, send you solution and proof of how we can solve. For demonstration, or design your best, write.

ElectroData

Division of International Corporation
with world-wide sales and service facilities
409 Sierra Madre Villa
Pasadena, California

35

Egy hirdetés 1957, amely a számítógépet óriási agyként ábrázolja.

A számítógépes programok gyorsan lehetővé tették, hogy a gépek megfeleljenek ennek az elvárásnak. Az 1959-ben Herbert A. Simon, J. C. Shaw és Allen Newell által a Carnegie Mellon Egyetemen létrehozott "általános problémamegoldó" képes volt kidolgozni

egy olyan tétel bizonyítását, amelyet Bertrand Russell (1872-1970) és Alfred North Whitehead (1861-1947) matematikusok 1913-ban írt híres *Principia Mathematica című* munkájukban nem tudtak megoldani. Az ezt követő évtizedekben nyilvánvalóvá vált, hogy a számítógépek könnyen képesek jelentősen meghaladni az emberi képességeket olyan intellektuális feladatokban, mint a matematikai problémák megoldása, a betegségek diagnosztizálása és a sakkozás, de nehezen tudnak irányítani egy cipőfűzőt kötő robotot vagy megérteni egy ötéves gyermek által megértett köznapi nyelvet. A számítógépek csak most kezdik elsajátítani ezeket a képességeket. Ironikus módon a számítógépes intelligencia fejlődése az emberi érettséggel ellentétes irányba haladt.

Az a kérdés, hogy a számítógép és az emberi agy valamilyen szinten egyenértékű-e vagy sem, ma is vitatott. A bevezetőben említettem, hogy az emberi agy komplexitásával kapcsolatos idézeteket milliónyi linkre lehetett hivatkozni. Hasonlóképpen, a Google "Idézetek: Az agy nem számítógép" keresése szintén több millió linket ad vissza. Véleményem szerint az ilyen irányú kijelentések olyanok, mintha azt mondanánk, hogy "az almászós nem alma". Technikailag ez az állítás igaz, de almászósz lehet almából készíteni. Talán még inkább olyan, mintha azt mondanánk, hogy "A számítógép nem szövegszerkesztő". Igaz, hogy a számítógép és a szövegszerkesztő különböző fogalmi szinteken létezik, de egy számítógép akkor válhat szövegszerkesztőgé, ha szövegszerkesztő szoftvert futtat, egyébként pedig nem. Hasonlóképpen, egy számítógép akkor válhat aggyá, ha agyszoftver fut rajta. Erre tesznek kísérletet a kutatók, köztük én is.

A kérdés tehát az, hogy találunk-e olyan algoritmust, amely a számítógépet az emberi aggyal egyenértékű entitássá alakítja. Egy számítógép ugyanis veleszületett univerzalitása miatt (csak a kapacitásától függően) bármilyen általunk definiált algoritmust képes futtatni. Az emberi agy viszont algoritmusok egy meghatározott halmazát futtatja. Módszerei okosak, mivel jelentős plaszticitást és saját kapcsolatainak átstrukturálását teszi lehetővé a tapasztalatai alapján, de ezek a funkciók szoftverben is emulálhatók.

A számítás egyetemessége (az az elképzelés, hogy egy általános célú számítógép bármilyen algoritmust képes végrehajtani) - és ennek az elképzelésnek a hatalma - az első tényleges gépekkel egy időben jelent meg. Négy kulcsfogalom áll a számítás egyetemességének és megvalósíthatóságának, valamint gondolkodásunkban való alkalmazhatóságának hátterében. Ezeket érdemes itt áttekinteni, mert az agy maga is é velük. Az első az információ megbízható közlésének, emlékezésének és kiszámításának képessége. Régebben, ha 1940, a "számítógép" szót használtuk, az emberek azt feltételezték, hogy analóg számítógépről beszélünk, amelyben a számokat különböző feszültség szintek reprezentálták, és a speciális alkatrészek olyan aritmetikai funkciókat tudtak végrehajtani, mint az összeadás és a szorzás. Az analóg számítógépek nagy korlátja azonban az volt, hogy pontossági problémák gyötörték őket. A számokat csak körülbelül százból egy rész pontossággal lehetett ábrázolni, és ahogy a számokat reprezentáló feszültség szinteket egyre több aritmetikai operátor dolgozta fel, a hibák felhalmozódtak. Ha egy maroknyi számításnál többet akartunk elvégezni, az eredmények annyira pontatlanná váltak, hogy értelmetlenné váltak.

Bárki, aki emlékszik az analóg szalagos magnókkal történő zenei felvételek idejére, emlékszik erre a hatásra. Az első példányon érezhető volt a romlás, mivel egy kicsit

zajosabb volt, mint az eredeti. (Ne feledjük, hogy a "zaj" véletlenszerű pontatlanságokat jelent.) A másolat másolata még zajosabb volt, és a tizedik generációnál a másolat már szinte teljesen zajos volt. Feltételezték, hogy ugyanez a probléma fogja sújtani a digitális számítógépek kialakulóban lévő világát is. Megérthetjük az ilyen aggodalmakat, ha a digitális információ csatornán keresztüli kommunikációját tekintjük. Egyetlen csatorna sem tökéletes, és minden csatornának van valamilyen eredendő hibaaránya. Tegyük fel, hogy van egy olyan csatornánk, amely 0,9 valószínűséggel helyesen továbbít minden egyes bitet. Ha egy egybites üzenetet küldök, akkor a csatornán keresztül történő pontos továbbítás valószínűsége a következő: Tegyük.9. fel, hogy két bitet küldök. Most a pontosság $.9^2 = .81$. Mi a helyzet, ha egy bájt (nyolc bit) küldök? Kevesebb, mint páros esélyem van (.43, hogy pontos legyenek) a helyes elküldésre. Öt bájt pontos elküldésének valószínűsége körülbelül 1 százalék.

Ennek a problémának a megkerülésére kézenfekvő megoldás a csatorna pontosabbá tétele. Tegyük fel, hogy a csatorna egymillió bitből csak egyet hibázik. Ha elküldök egy félmillió bájtból álló fájlt (ami körülbelül egy szerény program vagy adatbázis mérete), akkor a csatorna nagyon nagy eredendő pontossága ellenére a helyes továbbítás valószínűsége kevesebb mint 2 százalék. Tekintettel arra, hogy egyetlen bites hiba teljesen érvényteleníthet egy számítógépes programot és a digitális adatok más formáit, ez nem kielégítő helyzet. Függetlenül a csatorna pontosságától, mivel a hiba valószínűsége az átvitelben az üzenet méretével gyorsan nő, ez egy megoldhatatlan akadálnak tűnik.

Az analóg számítógépek ezt a problémát a kíméletes leépítésen keresztül közelítették meg (ami azt jelenti, hogy a felhasználók csak olyan problémákat mutattak be, amelyekben kis hibákat is el tudtak viselni); azonban ha az analóg számítógépek felhasználói a számítások korlátozott körére korlátozták magukat, a számítógépek némileg hasznosnak bizonyultak. A digitális számítógépek viszont folyamatos kommunikációt igényelnek, nemcsak az egyik számítógéptől a másikig, hanem magán a számítógépen belül is. A memóriából a központi feldolgozó egység felé és a központi feldolgozó egységből is van kommunikáció. A központi feldolgozó egységen belül kommunikáció folyik az egyik regisztertől a másikig, majd vissza és tovább az aritmetikai egységhez, és így tovább. Még az aritmetikai egységen belül is van kommunikáció az egyik bitregisztertől a másikig. A kommunikáció minden szinten mindenütt jelen van. Ha figyelembe vesszük, hogy a kommunikáció növekedésével a hibaarányok gyorsan növekednek, és hogy egyetlen bithiba is képes tönkretenni egy folyamat integritását, a digitális számítás halálra volt ítélve - legalábbis akkoriban úgy tűnt.

Figyelemre méltó módon ez volt az általános nézet, amíg Claude Shannon (1916-2001) amerikai matematikus meg nem jelent, és be nem mutatta, hogyan lehet tetszőlegesen pontos kommunikációt létrehozni még a legmegbízhatóbb kommunikációs csatornák segítségével is. Amit Shannon a *Bell System Technical Journal* című folyóiratban 1948 júliusában és októberében megjelent "A Mathematical Theory of Communication" című, mérföldkőnek számító tanulmányában, és különösen a zajos csatornakódolásra vonatkozó tételében kifejtett, az az volt, hogy ha rendelkezésünkre áll egy olyan csatorna, amelynek hibaaránya tetszőleges (kivéve pontosan 50 százalékos hibaarányt bitenként, ami azt

jelenténi, hogy a csatorna csak tiszta zajt közvetít), akkor képesek vagyunk olyan üzenetet továbbítani, amelyben a hibaarány olyan pontos, amelyet csak akarunk. Más szóval, az átvitel hibaaránya lehet egy bit az n bitből, ahol az n olyan nagy lehet, amilyennek Ön meghatározza. Így például szélsőséges esetben, ha van egy olyan csatornánk, amely az információ bitjeinek csak 51 százalékát továbbítja helyesen (azaz a helyes bitet csak valamivel gyakrabban továbbítja, mint a helytelen bitet), akkor ennek ellenére olyan üzeneteket továbbíthatunk, hogy egymillióból csak egy bit hibás, vagy egy bit a trillióból vagy trillió trillióból egy bit hibás.

Hogyan lehetséges ez? A válasz a redundancia. Ez ma már nyilvánvalónak tűnhet, de akkoriban nem volt az. Egy egyszerű példával élve, ha minden egyes bitet háromszor továbbítok, és a többségi szavazatot veszem, akkor jelentősen megnöveltem az eredmény megbízhatóságát. Ha ez nem elég jó, egyszerűen növelje a redundanciát, amíg el nem éri a szükséges megbízhatóságot. Az információ egyszerű megismétlése a legegyszerűbb módja annak, hogy alacsony pontosságú csatornákból tetszőlegesen magas pontossági arányokat érjünk el, de nem ez a leghatékonyabb megközelítés. Shannon tanulmánya, amely megalapozta az információelméletet, olyan optimális hibadetektálási és hibajavító kódokat mutatott be, amelyekkel *bármilyen* nem véletlenszerű csatornán keresztül *bármilyen* célpontosságot el lehet érni.

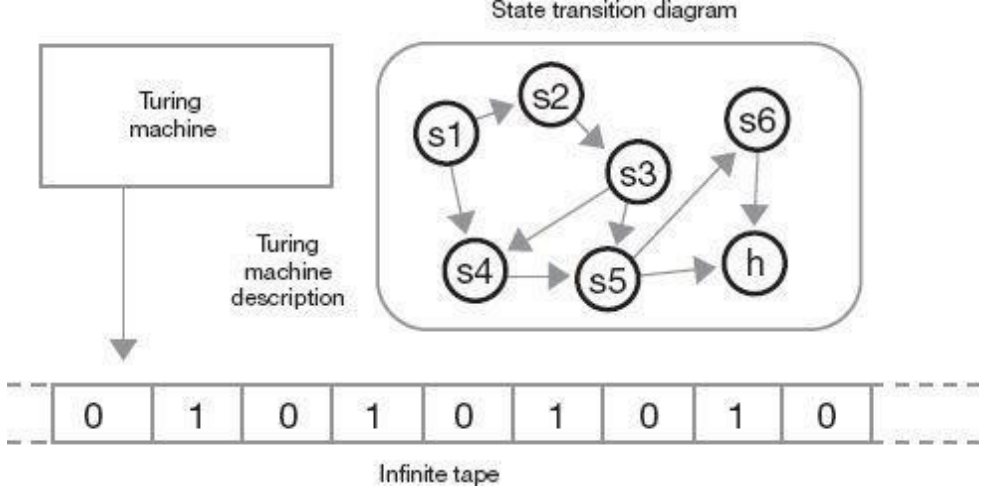
Az idősebb olvasók emlékeznek a telefonmodemekre, amelyek zajos analóg telefonvonalakon keresztül továbbították az információkat. Ezek a vonalak hallhatóan sziszegést, pukkanást és a torzítás számos más formáját mutatták, de ennek ellenére képesek voltak a digitális adatok igen nagy pontosságú továbbítására, köszönhetően Shannon zajos csatorna-tételének. Ugyanez a probléma és ugyanez a megoldás létezik a digitális memória esetében is. Gondolkodott már azon, hogy a CD-k, DVD-k és programlemezek hogyan szolgáltatnak továbbra is megbízható eredményeket még akkor is, ha a lemezt a padlóra ejtették és megkarcolták? Ismét Shannonnak köszönhetjük.

A számítás három elemből áll: kommunikáció - amely, mint említettem, mind a számítógépeken belül, mind a számítógépek között mindenütt jelen van -, memória, és logikai kapuk (amelyek az aritmetikai és logikai funkciókat végzik). A logikai kapuk pontossága tetszőlegesen nagyra növelhető a hibadetektáló és -javító kódok hasonló alkalmazásával. Shannon tételének és elméletének köszönhető, hogy tetszőlegesen nagy és összetett digitális adatokat és algoritmusokat tudunk kezelni anélkül, hogy a folyamatokat megzavarnák vagy tönkretennék a hibák. Fontos kiemelni, hogy az agy is Shannon elvét használja, bár az emberi agy evolúciója egyértelműen megelőzte Shannont! A legtöbb mintát vagy ötletet (és egy ötlet is egy minta), mint láttuk, az agy jelentős redundanciával tárolja. Az agyban lévő redundancia elsődleges oka az idegi áramkörök eredendő megbízhatatlansága.

A második fontos gondolat, amelyre az információs korszak támaszkodik, az, amit korábban említettem: a számítás egyetemessége. Alan Turing 1936-ban írta le "Turing-gépét", amely nem egy valódi gép, hanem egy újabb gondolat kísérlet volt. Elméleti számítógépe egy végtelen hosszú memóriaszalagból áll, amelynek minden négyzetében egy 1 vagy egy 0 van. A gép bemenete ezen a szalagon jelenik meg, amelyet a gép egyszerre csak egy-egy négyzetet tud beolvasni. A gép tartalmaz egy szabálytáblázatot is - lényegében

egy tárolt programot -, amely számozott állapotokból áll. Minden szabály egy műveletet határoz meg, ha az éppen olvasott négyzet egy, és 0, egy másik műveletet, ha az aktuális négyzet 1. A lehetséges műveletek közé tartozik a 0 vagy 1 írása a szalagra, a szalag egy négyzettel jobbra vagy balra mozgatása, vagy a megállás. Minden állapot ezután megadja a következő állapot számát, amelyben a gépnek lennie kell. A Turing-gép bemenete a szalagon jelenik meg. A program lefut, és amikor a gép megáll, befejezte az algoritmusát, és a folyamat kimenete a szalagon marad. Vegyük észre, hogy bár a szalag elméletileg végtelen hosszúságú, bármely tényleges program, amely nem kerül végtelen ciklusba, a szalagnak csak egy véges részét használja fel, így ha véges szalagra szorítkozunk, a gép még mindig hasznos problémákat fog megoldani.

Ha a Turing-gép egyszerűnek hangzik, az azért van, mert ez volt a feltalálójának célja. Turing azt akarta, hogy a gépe a lehető legegyszerűbb legyen (de Einsteint idézve nem egyszerűbb). Turing és Alonzo Church (1903-1995), egykori professzora közösen dolgozták ki a Church-Turing tézist, amely szerint ha egy Turing-gép elé állítható probléma nem oldható meg a Turing-gép által, akkor a természettörvényt követve *semmilyen* gép által sem oldható meg. Annak ellenére, hogy a Turing-gép csak néhány paranccsal rendelkezik, és egyszerre csak egy bitet dolgoz fel, mégis mindent ki tud számolni, amit bármely számítógép ki tud számolni. Másképpen úgy is mondhatjuk, hogy bármelyik "Turing-teljes" (azaz a Turing-gépével egyenértékű képességekkel rendelkező) gép képes bármilyen algoritmust (bármely általunk definiálható eljárást) kiszámítani.



Egy Turing-gép blokkdiagramja, amely egy fejvel rendelkezik, amely a szalagot olvassa és írja, valamint egy belső programmal, amely állapotátmenetekből áll.

A Church-Turing tézis "erős" értelmezései alapvető egyenértékűséget javasolnak aközött, amit egy ember gondolhat vagy tudhat, és ami egy gép által kiszámítható. Az alapgondolat az, hogy az emberi agy szintén a természeti törvényeknek van alávetve, és így információfeldolgozó képessége nem haladhatja meg egy gép (és így egy Turing-gép) képességét.

Turingnak tulajdoníthatjuk, hogy 1936-os tanulmányával megteremtette a számítás elméleti alapjait, de fontos megjegyezni, hogy nagy hatással volt rá egy előadás, amelyet John von Neumann (1903-1957) amerikai magyar matematikus tartott Cambridge-ben 1935a tárolt program koncepciójáról, egy olyan koncepcióról, amely a számítás elméleti alapjait teremtette meg.

a Turing-gépben rejlik.¹ Von Neumannra viszont hatással volt Turing 1936-os tanulmánya, amely elegánsan lefektette a számítás alapelveit, és az 1930-as évek végén és az 1940-es évek elején kötelező olvasmánnyá tette kollégái számára.²

Ugyanebben a tanulmányban Turing egy másik váratlan felfedezésről is beszámol: a megoldhatatlan problémákról. Ezek olyan problémák, amelyek jól definiáltak, egyedi válaszokkal, amelyekről bebizonyítható, hogy léteznek, de amelyekről azt is be tudjuk bizonyítani, hogy soha semmilyen Turing-gép nem tudja kiszámítani őket - vagyis *semmilyen* gép nem tudja kiszámítani őket, ami megfordítja azt a XIX. századi dogmát, hogy a definiálható problémák végül is megoldhatók. Turing kimutatta, hogy ugyanannyi megoldhatatlan probléma létezik, mint megoldható. Kurt Gödel osztrák származású amerikai matematikus és filozófus 1931-es "befejezetlenségi tételében" hasonló következtetésre jutott. Így maradunk abban a zavarba ejtő helyzetben, hogy képesek vagyunk meghatározni egy problémát, bebizonyítani, hogy létezik egy egyedi válasz, és mégis tudjuk, hogy a választ soha nem lehet megtalálni.

Turing megmutatta, hogy a számítás lényegét tekintve egy nagyon egyszerű mechanizmuson alapul. Mivel a Turing-gép (és így bármely számítógép) képes arra, hogy

jövőbeli cselekvéseit a már kiszámított eredményekre alapozza, képes döntéseket hozni és tetszőlegesen összetett információhierarchiákat modellezni.

1939-ben Turing megtervezte a Bombe nevű elektronikus számológépet, amely segítette a náci Enigma kódológép által titkosított üzenetek dekódolásában. Egy 1943, Turing által befolyásolt mérnöksapat elkészítette a vitathatatlanul első számítógépet, a Colossust, amely lehetővé tette a szövetségesek számára, hogy folytassák az Enigma kifinomultabb változataiból származó üzenetek dekódolását. A Bombe és a Colossus egyetlen feladatra készült, és nem lehetett őket más feladatra átprogramozni. Ezt a feladatot azonban kiválóan teljesítették, és nekik tulajdonítják, hogy lehetővé tették a szövetségesek számára, hogy legyőzzék a német Luftwaffe három az egyhez arányú előnyét a brit Királyi Légierővel szemben, és megnyerjék a döntő fontosságú angliai csatát, valamint hogy a háború során folyamatosan megelőzzék a náci taktikákat.

John von Neumann ezen az alapon alkotta meg a modern számítógép architektúráját, amely a harmadik nagy ötletünket képviseli. A von Neumann-gépnek nevezett architektúra az elmúlt hatvanhét évben lényegében minden számítógép alapvető szerkezete maradt, a mosógépben lévő mikrokontrollertől a legnagyobb szuperszámítógépekig. 1945. június 30-án kelt, "First Draft of a Report on the EDVAC" (Az EDVAC-ról szóló jelentés első tervezete) című tanulmányában von Neumann bemutatta azokat az elképzeléseket, amelyek azóta is uralják a számítást.³ A von Neumann-modell tartalmaz egy központi feldolgozóegységet, ahol aritmetikai és logikai műveleteket végeznek; egy memóriaegységet,

ahol a programot és az adatokat tárolják; tömegtároló; programszámláló; és bemeneti/kimeneti csatornák. Bár ezt a dokumentumot belső projektdokumentumnak szánták, mára a számítógéptervezők bibliájává vált. Soha nem lehet tudni, hogy egy látszólag rutinszerű belső feljegyzés végül forradalmasítja a világot.

A Turing-gépet nem a gyakorlati alkalmazhatóságra tervezték. Turing tételei nem a problémák megoldásának hatékonyságával foglalkoztak, hanem inkább a problémák azon körét vizsgálták, amelyek elméletileg számítással megoldhatók. Von Neumann célja ezzel szemben a számítógép megvalósítható koncepciójának megalkotása volt. Az ő modellje Turing egybites számításait többbites szavakkal (általában a nyolc bit valamilyen többszörösével) helyettesíti. Turing memóriaszalagja szekvenciális, így a Turing-gép programjai a szalag ide-oda mozgatásával aránytalanul sok időt töltenek a köztes eredmények tárolására és előhívására. Ezzel szemben a von Neumann-memória véletlen hozzáférésű, így bármely adatelem azonnal előhívható.

Von Neumann egyik legfontosabb ötlete a tárolt program, amelyet egy évtizeddel korábban vezetett be: a programot ugyanolyan típusú véletlen hozzáférésű memóriában helyezi el, mint az adatokat (és gyakran ugyanabban a memóriablokkban). Ez lehetővé teszi a számítógép átprogramozását különböző feladatokra, valamint az önmódosító kódot (ha a programtároló írható), ami a rekurzió egy erőteljes formáját teszi lehetővé. Addig gyakorlatilag minden számítógép, így a Colossus is, egy adott feladatra készült. A tárolt program lehetővé teszi, hogy egy számítógép valóban univerzális legyen, és ezzel teljesül Turing elképzelése a számítás univerzalitásáról.

A von Neumann-gép másik kulcsfontosságú jellemzője, hogy minden utasítás tartalmaz egy műveleti kódot, amely meghatározza az elvégzendő aritmetikai vagy logikai műveletet és a memóriából származó operandus címét.

Von Neumann koncepciója arról, hogyan kell egy számítógépet felépíteni, az EDVAC tervének közzétételével jelent meg, amelyet J. Presper Eckert és John Mauchly munkatársaival közösen végzett. Maga az EDVAC 1951-ig nem is működött, ekkorra már léteztek más tárolt programú számítógépek, mint például a Manchester Small-Scale Experimental Machine, az ENIAC, az EDSAC és a BINAC, amelyek mindegyikére nagy hatással volt von Neumann tanulmánya, és amelyek tervezői Eckert és Mauchly voltak. Von Neumann közvetlenül közreműködött számos ilyen gép tervezésében, beleértve az ENIAC későbbi változatát, amely támogatta a tárolt programot.

A von Neumann-architektúrának volt néhány előfutára, bár egy meglepő kivételtől eltekintve egyik sem igazi von Neumann-gép. Howard Aiken 1944-ben mutatta be a Mark I-et, amely rendelkezett a programozhatóság egy elemével, de nem használt tárolt programot. Az utasításokat egy lyukasztott papírszalagról olvasta be, majd minden egyes parancsot azonnal végrehajtott. Nem rendelkezett feltételes elágazási utasítással sem.

Konrad Zuse (1910-1995) német 1941 tudós megalkotta a Z-3 számítógépet. Ez is szalagról olvasta a programját (ebben az esetben filmre kódolva), és nem rendelkezett feltételes elágazási utasítással. Érdekes, hogy Zuse támogatást kapott a Német Repülőgépkutató Intézettől, amely a készüléket a szárnyrepülések tanulmányozására használta, de a náci kormánynak benyújtott, a reléinek vákuumcsövekkel való helyettesítésére irányuló finanszírozási javaslatát elutasították. A nácik a számítást "nem háborús szempontból fontosnak" tartották. Véleményem szerint ez a nézőpont nagyban hozzájárul a háború kimenetelének magyarázatához.

Von Neumann koncepciójának valójában van egy valódi előfutára, és ez egy teljes évszázaddal korábbi! Charles Babbage (1791-1871) angol matematikus és feltaláló analitikus motorja, amelyet először 1837-ben írt le, valóban magába foglalta von Neumann ötleteit, és egy lyukasztón keresztül tárolt programot tartalmazott.

a Jacquard szövőszékről kölcsönzött kártyák.⁴ A véletlen hozzáférésű memória 1000, egyenként 50 tizedesjegű szót tartalmazott (ami körülbelül 21 kilobájtak felel meg). Minden utasítás tartalmazott egy op-kódot és egy operandusszámot, akárcsak a modern gépi nyelvek. Tartalmazta a feltételes elágazást és a hurkolás funkciót, tehát igazi von Neumann-gép volt. Teljesen mechanikus fogaskerekeken alapult, és úgy tűnik, hogy az Analytical Engine meghaladta Babbage tervezési és szervezési képességeit. Részzeit megépítette, de soha nem futott. Nem világos, hogy a számítógép huszadik századi úttörői, beleértve von Neumann-t is, tudtak-e Babbage munkájáról.

Babbage számítógépe a szoftverprogramozás területének megteremtését eredményezte. Ada Byron (1815-1852) angol író, Lovelace grófnő, Lord Byron költő egyetlen törvényes gyermeke volt a világ első számítógépes programozója. Programokat írt az Analytical Engine számára, amelyeket saját fejében kellett hibakereséssel javítania (mivel a számítógép soha nem működött), ezt a gyakorlatot ma a szoftvermérnökök jól ismerik "táblaellenőrzés" néven. Lefordította Luigi Menabrea olasz matematikusnak az Analytical

Engine-ről szóló cikkét, és részletes saját jegyzeteket fűzött hozzá, amelyekben azt írta, hogy "az Analytical Engine algebrai mintákat sző, ahogy a Jacquard szövőszék virágokat és leveleket sző". A továbbiakban talán az első találgatások a mesterséges intelligencia megvalósíthatóságáról, de arra a következtetésre jutott, hogy az Analytical Engine "semmiféle igényt nem támaszt arra, hogy bármit is létrehozzon".

Babbage koncepciója egészen csodálatos, ha figyelembe vesszük a korszakot, amelyben élt és dolgozott. A huszadik század közepére azonban elképzelései az idő ködébe vesznek (bár később újra felfedezték őket). Von Neumann volt az, aki megfogalmazta és megfogalmazta a ma ismert számítógép legfontosabb alapelveit, és a világ ezt azzal ismeri el, hogy továbbra is a von Neumann-gépre mint a számítás fő modelljére hivatkozik. Ne feledjük azonban, hogy a von Neumann-gép folyamatosan adatokat közöl a különböző egységei között és ezeken az egységeken belül, így nem lehetett volna megépíteni Shannon tételei és az általa a megbízható digitális információ továbbítására és tárolására kidolgozott módszerek nélkül.

Ezzel elérkeztünk a negyedik fontos gondolathoz, amely túlmutat Ada Byron következtetésén, miszerint egy számítógép nem képes kreatívan gondolkodni, és meg kell találnunk az agy által használt kulcsfontosságú algoritmusokat, majd ezeket felhasználva a számítógépet agyakká kell alakítanunk. Alan Turing ezt a célt a "Computing Machinery and Intelligence" 1950 (Számítógépek és intelligencia) című tanulmányában mutatta be, amely tartalmazza a ma már híres Turing-tesztjét annak megállapítására, hogy egy mesterséges intelligencia elérte-e az emberi intelligencia szintjét.

1956-ban von Neumann elkezdett egy előadásorozatot előkészíteni, amelyet a Yale Egyetem rangos Silliman előadásorozatára szánt. A rákos megbetegedések miatt ezeket az előadásokat soha nem tartotta meg, és nem fejezte be a kéziratot sem, amelyből azokat előadta volna. Ez a befejezetlen dokumentum mindazonáltal ragyogó és prófétai előképe maradt annak, amit én az emberiség legnehezebb és legfontosabb projektjének tartok. Posztumusz jelent meg 1958-ban *The Computer and the Brain (A számítógép és az agy)* címmel. Találó, hogy a múlt század egyik legzseniálisabb matematikusának és a számítógépes korszak egyik úttörőjének utolsó munkája magának az intelligenciának a vizsgálata volt. Ez a projekt volt a legkorábbi komoly vizsgálat az emberi agyról egy matematikus és informatikus szemszögéből. Von Neumann előtt a számítástechnika és az idegtudomány két sziget volt, amelyek között nem volt híd.

Von Neumann a számítógép és az emberi agy közötti hasonlóságok és különbségek megfogalmazásával kezdi a vitát. Tekintettel arra, hogy mikor írta ezt a kéziratot, ez figyelemre méltóan pontos. Megjegyezte, hogy a neuronok kimenete digitális - egy axon vagy tüzel, vagy nem. Ez akkoriban messze nem volt nyilvánvaló, hiszen a kimenet lehetett volna analóg jel is. A neuronba vezető dendritekben és a szoma neuron sejttestében zajló feldolgozás azonban analóg volt, és számításait a bemenetek küszöbértékkel súlyozott összegeként írta le. A neuronok működésének ez a modellje vezetett a konnekcionizmus területéhez, amely hardverben és szoftverben is erre a neuronmodellre épülő rendszereket épített. (Amint azt az [előző fejezetben](#) leírtam, az első ilyen konnekcionista rendszert Frank Rosenblatt hozta létre szoftverprogramként egy IBM számítógépen 704 a Cornellben 1957-ben, közvetlenül azután, hogy von Neumann előadásainak vázlata elérhetővé vált). Ma már

kifinomultabb modellekkkel rendelkezőnk arról, hogy a neuronok hogyan kombinálják a bemeneteket, de a dendritbemenetek analóg feldolgozásának alapvető gondolata a neurotranszmitter-koncentrációk segítségével érvényes maradt.

Von Neumann a számítás egyetemességének koncepcióját alkalmazva arra a következtetésre jutott, hogy bár az agy és a számítógép felépítése és építőkövei radikálisan különböznek egymástól, mégis megállapíthatjuk, hogy egy von Neumann-gép képes szimulálni az agyban zajló feldolgozást. A fordítottja azonban nem igaz, mert az agy nem von Neumann-gép, és nem rendelkezik tárolt programmal mint olyannal (bár egy nagyon egyszerű Turing-gépet tudunk szimulálni a fejünkben). Az algoritmusai vagy módszerei implicit módon benne vannak a szerkezetében. Von Neumann helyesen következtet arra, hogy a neuronok képesek mintákat tanulni a bemeneteikből, amelyekről ma már tudjuk, hogy részben a dendritek erősségében vannak kódolva. Amit von Neumann idejében még nem tudtak, az az, hogy a tanulás a neuronok közötti kapcsolatok létrehozásán és megsemmisítésén keresztül is zajlik. Von Neumann előrelátóan megjegyzi, hogy az idegi feldolgozás sebessége rendkívül lassú, másodpercenként száz számítás nagyságrendű, de az agy ezt masszív párhuzamos feldolgozással kompenzálja - ez egy másik nem nyilvánvaló és kulcsfontosságú felismerés. Von Neumann azt állította, hogy az agy 10^{10} neuronjának mindegyike (ez a számadat önmagában is meglehetősen pontos volt; a mai becslések szerint 10^9 és 10^{11} között van) egyszerre dolgozik. Valójában minden egyes összeköttetés (neurononként átlagosan körülbelül 10^3 összeköttetéssel) egyszerre számol. Von Neumann becslései és a neurális feldolgozás leírása figyelemre méltó, figyelembe véve az idegtudomány akkori kezdetleges állapotát. Munkájának egy olyan aspektusa azonban, amellyel nem értek egyet, az agy memóriakapacitására vonatkozó becslése. Feltételezi, hogy az agy egész életén át emlékszik minden bemenetre. Von Neumann 60 éves átlagos élettartamot feltételez, ami körülbelül 2×10^9 másodpercet jelent. Az egyes neuronokhoz másodpercenként körülbelül 14 bemenettel (ami valójában legalább három nagyságrenddel alacsonyabb) és neuronokkal 10^{10} számolva az agy memóriakapacitását körülbelül 10^{20} bite becsüli. A valóság az, ahogyan azt már korábban megjegyeztem, hogy gondolatainknak és tapasztalatainknak csak nagyon kis hányadára emlékszünk, és még ezek az emlékek sem alacsony szintű bitminták formájában (mint például egy videokép), hanem magasabb szintű minták sorozataként tárolódnak. Ahogy von Neumann leírja az egyes agyi mechanizmusokat, megmutatja, hogy egy modern számítógép a látszólagos különbségek ellenére hogyan tudná ugyanazt megvalósítani. Az agy analóg mechanizmusai digitális mechanizmusokkal szimulálhatók, mivel a digitális számítás bármilyen kívánt pontossággal képes az analóg értékek utánzására (és az agyban az analóg információ pontossága meglehetősen alacsony). Az agy masszív párhuzamossága is szimulálható, mivel a számítógépek jelentős sebességelőnye a soros számításban rejlik (ez az előny az idők során jelentősen megnőtt). Emellett a számítógépekben is használhatunk párhuzamos feldolgozást, párhuzamos von Neumann-gépek segítségével - pontosan így működnek ma a szuperszámítógépek.

Von Neumann arra a következtetésre jut, hogy az agy módszerei nem tartalmazhatnak hosszadalmas, szekvenciális algoritmusokat, ha figyelembe vesszük, hogy az emberek milyen gyorsan képesek döntéseket hozni, miközben a neuronok számítási sebessége nagyon

lassú. Amikor egy harmadik bázisjátékos labdát szerez, és úgy dönt, hogy a második bázis helyett az elsőre dob, ezt a döntést a másodperc töredéke alatt hozza meg, ami minden neuron számára csupán néhány ciklusra elegendő idő. Von Neumann helyesen következtet arra, hogy az agy figyelemre méltó képességei abból erednek, hogy mind a 100 milliárd neuron egyidejűleg képes feldolgozni az információt. Amint azt már megjegyeztem, a látókéreg mindössze három-négy idegi ciklus alatt bonyolult vizuális ítéleteket hoz.

Az agyban jelentős plaszticitás van, ami lehetővé teszi számunkra a tanulást. De sokkal nagyobb a plaszticitás egy számítógépben, amely a szoftverének megváltoztatásával teljesen át tudja alakítani a módszereit. Ebben a tekintetben tehát a számítógép képes lesz utánozni az agyat, de ez fordítva nem igaz.

Amikor von Neumann összehasonlította az agy tömegesen párhuzamos szerveződésének kapacitását a korabeli (kevés) számítógéppel, egyértelművé vált, hogy az agy sokkal nagyobb memóriával és sebességgel rendelkezik. Mostanra az első olyan szuperszámítógép, amely elérte a funkcionális működéshez szükséges sebességre vonatkozó óvatosabb becsléseknek megfelelő specifikációkat, már nem volt olyan nagy.

az emberi agyat szimuláló (másodpercenként kb. 10^{16} művelet) számítógépet építettek.⁵ (Becsléseim szerint a 2020-as évek elejére 1000 dollárba fog kerülni ez a szintű számítás.) A memória tekintetében még közelebb vagyunk. Bár a kézirat megírásakor még feltűnően korán volt a számítógép történetében, von Neumann mégis bízott abban, hogy az emberi intelligencia hardvere és szoftvere is a helyére kerül végül, és ez volt a motivációja arra, hogy elkészítse ezeket az előadásokat.

Von Neumann mélyen tisztában volt a fejlődés egyre gyorsuló ütemével és annak az emberiség jövőjére gyakorolt mélyreható következményeivel. Egy évvel halála után, Stan Ulam 1957, matematikus kollégája idézte őt, aki az 1950-es évek elején azt mondta, hogy "a technológia egyre gyorsuló fejlődése és az emberi életmódban bekövetkező változások azt a látszatot keltik, hogy a faj történetében közeledik egy olyan alapvető szingularitás, amelyen túl az emberi ügyek, ahogyan mi ismerjük őket, nem folytatódhatnak". Ez a "szingularitás" szó első ismert használata az emberi technikatörténettel összefüggésben.

Von Neumann alapvető felismerése az volt, hogy a számítógép és az agy között alapvető egyenértékűség van. Vegyük észre, hogy a biológiai ember érzelmi intelligenciája része az intelligenciájának. Ha von Neumann felismerése helyes, és ha elfogadjuk a saját hitemet, hogy egy nem biológiai entitás, amely meggyőzően újra létrehozza egy biológiai ember (érzelmi és egyéb) intelligenciáját, tudatos (lásd a [következő fejezetet](#)), akkor azt a következtetést kell levonnunk, hogy a számítógép és az agy között alapvető egyenértékűség van.

-a megfelelő szoftverrel - és (tudatos) elmével. Tehát von Neumann-nak igaza van?

A legtöbb számítógép ma már teljesen digitális, míg az emberi agy a digitális és az analóg módszereket ötvözi. Az analóg módszerek azonban könnyen és rutinszerűen újraalkothatók digitális módszerekkel, bármilyen kívánt pontossággal. Carver Mead (1934-ben született) amerikai informatikus megmutatta, hogy az agy analóg módszereit közvetlenül emulálhatjuk szilíciumban, amit az általa "neuromorfikusnak" nevezett chipekkel demonstrált.⁶ Mead bemutatta, hogy ez a megközelítés több ezerszer hatékonyabb lehet,

mint az analóg módszerek digitális emulálása. A tömegesen ismétlődő neokortikális algoritmus kodifikálásakor érdemes lesz Mead megközelítését alkalmazni. A Dharmendra Modha által vezetett IBM Cognitive Computing Group olyan chipeket mutatott be, amelyek neuronokat és azok kapcsolatait emulálják, beleértve az új kapcsolatok kialakításának képességét is.⁷ A "SyNAPSE" nevű chipek egyike 256 neuron közvetlen szimulációját biztosítja, mintegy negyedmillió szinaptikus kapcsolattal. A projekt célja egy 10 milliárd neuronnal és 100 trillió kapcsolattal rendelkező - az emberi agyhoz közeli - szimulált neokortex létrehozása, amely mindössze egy kilowatt energiát használ.

Ahogy von Neumann több mint fél évszázaddal ezelőtt leírta, az agy rendkívül lassú, de rendkívül párhuzamos. A mai digitális áramkörök legalább 10 milliószor gyorsabbak, mint az agy elektrokémiai kapcsolói. Ezzel szemben az agy mind a 300 millió neokortikális mintafelismerője egyszerre dolgozik, és az agy mind a négy milliárd interneuronális kapcsolata potenciálisan egyszerre számol. Az emberi agy sikeres modellezéséhez szükséges hardver biztosításának kulcskérdése azonban a szükséges teljes memória és számítási teljesítmény. Nem kell közvetlenül lemásolnunk az agy architektúráját, ami nagyon gazdaságtalan és rugalmatlan megközelítés lenne.

Becsüljük meg, hogy mik ezek a hardverkövetelmények. Számos projekt próbálta meg utánozni azt a fajta hierarchikus tanulást és mintafelismerést, amely a neokortikális hierarchiában zajlik, beleértve a saját munkámat is, amely hierarchikus rejtett Markov-modellekkel foglalkozik. Saját tapasztalataim alapján egy óvatos becslés szerint a biológiai agy neokortexében lévő egyetlen mintafelismerő ciklus emulálása körülbelül számításokat 3,000 igényel. A legtöbb szimuláció ennek a becslésnek a töredékéért futnak. Ha az agy körülbelül $(10^2 \cdot 100)$ ciklus/másodperc sebességgel működik, akkor ez másodpercenként 3×10^5 (300 000) számítást jelent mintafelismerőnként. Ha az én becslésem szerint 3×10^8 (300 millió) mintafelismerővel számolunk, akkor körülbelül 10^{14} (100 billió) számítást kapunk másodpercenként, ami megegyezik a *The Singularity Is Near (A szingularitás közel van)* című könyvemben szereplő becsléssel. Abban a könyvben azt vetítettem előre, hogy az agy funkcionális szimulációjához másodpercenként 10 és 10^{14} közötti ¹⁶számításokra lenne szükség, és a 10^{16} cps értéket konzervatívnak tekintettem. Hans Moravec mesterséges intelligencia szakértő becslése, amely a korai (kezdeti) vizuális feldolgozás számítási igényének az egész agyra való extrapolálásán alapul, 10^{14} cps, ami megegyezik az én becsléssel.

A rutin asztali gépek elérhetik a 10^{10} cps sebességet, bár ez a teljesítményszint jelentősen növelhető a felhőalapú erőforrások használatával. A leggyorsabb szuperszámítógép, a japán K Computer már elérte a 10^{16} cps-ot.⁸ Tekintettel arra, hogy a neokortex algoritmusai tömegesen ismétlődnek, a neuromorf chipek, például a fent említett IBM SyNAPSE chipek használatának megközelítése is ígéretes.

A memóriaigényt tekintve körülbelül 30 bite (kb. négy bájt) van szükségünk egy kapcsolathoz, hogy 300 millió másik mintafelismerő közül egyet megszólítsunk. Ha úgy becsüljük, hogy átlagosan nyolc bemenet jut minden egyes mintafelismerőhöz, akkor ez felismerőnként bájtot 32 jelent. Ha minden bemenethez hozzáadunk egy egybájtos súlyt, akkor ez 40 bájtot jelent. Adjunk hozzá még 32 bájtot a lefelé irányuló kapcsolatokhoz, és

máris elérjük a bájtkokat. 72 Vegyük észre, hogy a felfelé és lefelé elágazó számadat gyakran sokkal magasabb lesz nyolcnál, bár ezek a nagyon nagy elágazó fák sok felismerő számára közősek. Például a "p" betű felismerésében több száz felismerő is részt vehet. Ezek több ezer ilyen felismerőbe táplálkoznak ezen a következő magasabb szinten, amelyek a "p" betűt tartalmazó szavakkal és kifejezésekkel foglalkoznak. Az egyes "p" felismerők azonban nem ismétlik meg azt a kapcsolati fát, amely a "p"-t tartalmazó összes szóhoz és kifejezéshez vezet - mindannyian egy ilyen kapcsolati fán osztoznak. Ugyanez igaz a lefelé irányuló kapcsolatokra is: Egy felismerő, amely az "APPLE" szóért felelős, az alatta lévő ezernyi "E" felismerőnek megmondja, hogy egy "E" várható, ha már látta az "A"-t, a "P"-t, a "P"-t és az "L"-et. Ez a kapcsolati fa nem ismétlődik meg minden olyan szó- vagy kifejezésfelismerő esetében, amely a következő, alacsonyabb szinten lévő felismerőt arról akarja tájékoztatni, hogy egy "E" várható. Ismétlem, ezek meg vannak osztva. Emiatt ésszerű az az általános becslés, hogy egy mintafelismerőre átlagosan nyolc felfelé és nyolc lefelé irányuló kapcsolat jut. Még ha növeljük is ezt a konkrét becslést, az nem változtatja meg jelentősen a kapott becslés nagyságrendjét.

Egyenként bájtnyi $72 \times (310^8 300 \text{ millió})$ mintafelismerővel számolva a teljes memóriaigény körülbelül $2 \times (10^{10} 20 \text{ milliárd})$ bájt. Ez valójában egy elég szerény szám, amelyet a mai rutinszámítógépek is képesek túllépni.

Ezek a becslések csak a szükséges nagyságrendek durva becslésére szolgálnak. Mivel a digitális áramkörök természetüknél fogva körülbelül 10 milliószor gyorsabbak, mint a biológiai neokortikális áramkörök, nem kell az emberi agy párhuzamosságát elérnünk - az emberi agy trilliószoros párhuzamosságához képest a legkevesebb párhuzamos feldolgozás is elegendő lesz. Láthatjuk, hogy a szükséges számítási követelmények elérhető közelségbe kerülnek. Az agy önmaga újrakábelezése - a dendritek folyamatosan új szinapszisokat hoznak létre - szintén emulálható szoftverben, linkek segítségével, ami sokkal rugalmasabb rendszer, mint az agy plaszticitási módszere, amely, mint láttuk, lenyűgöző, de korlátozott.

Az agy által a robusztus, változatlan eredmények eléréséhez használt redundancia minden bizonnyal megismételhető a szoftveres emulációkban. Az ilyen típusú önszerveződő hierarchikus tanuló rendszerek optimalizálásának matematikája jól ismert. Az agy szerveződése messze nem optimális. Persze nem is kellett annak lennie - csak elég jónak kellett lennie ahhoz, hogy elérje azt a küszöböt, hogy képes legyen olyan eszközöket létrehozni, amelyek kompenzálják a saját korlátait.

Az emberi neokortex másik korlátja, hogy nincs olyan folyamat, amely kiküszöbölné vagy akár csak felülvizsgálná az ellentmondásos gondolatokat, ami megmagyarázza, hogy az emberi gondolkodás gyakran miért olyan masszívan következtelen. Van egy gyenge mechanizmusunk ennek kezelésére, amit kritikus gondolkodásnak nevezünk, de ezt a képességet közel sem gyakoroljuk olyan gyakran, mint kellene. Egy szoftveralapú neokortexbe beépíthetünk egy olyan folyamatot, amely további felülvizsgálat céljából feltárja az ellentmondásokat.

Fontos megjegyezni, hogy egy teljes agyi régió tervezése egyszerűbb, mint egyetlen neuroné. Amint azt korábban már említettük, a modellek gyakran egyszerűbbé válnak magasabb szinten - gondoljunk csak a számítógép analógiájára. Egy tranzisztor

modellezéséhez azonban meg kell értenünk a félvezetők részletes fizikáját, és az egyetlen valódi tranzisztor alapjául szolgáló egyenletek összetettek. Egy digitális áramkör, amely két számot szoroz meg, több százat igényel. Mégis ezt a szorzási áramkört nagyon egyszerűen, egy-két képlettel tudjuk modellezni. Egy teljes számítógépet több milliárd tranziszttal modellezhetünk az utasításkészlet és a regiszterek leírásával, ami egy maroknyi írott oldalnyi szöveggel és képletekkel leírható. Egy operációs rendszer, a nyelvi fordítóprogramok és az assemblerek szoftverprogramjai meglehetősen összetettek, de egy adott program - például egy hierarchikus rejtett Markov-modellezésen alapuló beszédfelismerő program - modellezése szintén csak néhány oldalnyi egyenletben írható le. Egy ilyen leírásban sehol sem találunk a félvezetőfizika vagy akár a számítógép-architektúra részleteit.

Hasonló megfigyelés igaz az agyra is. Egy adott neokortikális mintafelismerő, amely egy adott invariáns vizuális jellemzőt (például egy arcot) észlel, vagy amely sávszűrést végez (a bemenetet egy adott frekvenciatartományra korlátozza) a hangon, vagy amely két esemény időbeli közelségét értékeli, sokkal kevesebb konkrét részlettel írható le, mint a neurotranszmittereket, ioncsatornákat és más, az idegi folyamatokban részt vevő szinaptikus és dendritikus változókat irányító tényleges fizikai és kémiai összefüggések.

Bár mindezt a komplexitást alaposan át kell gondolni, mielőtt a következő magasabb fogalmi szintre lépünk, az agy működési elveinek feltárásával sok minden leegyszerűsíthető.

FEJEZET 9

GONDOLATKÍSÉRLETEK AZ ELMÉVEL KAPCSOLATBAN

Az elme egyszerűen az, amit az agy csinál.

-Marvin Minsky, *Az elme társadalma*

Amikor intelligens gépeket építenek, nem kellene meglepődnünk azon, hogy ugyanolyan zavarosnak és makacsnak találjuk őket, mint az embereket az elme- anyaggal, a tudattal, a szabad akarattal és hasonlókkal kapcsolatos meggyőződésükben.

-Marvin Minsky, *Az elme társadalma* Ki a tudatos?

A tudat igazi története az ember első hazugságával kezdődik.

-Joseph Brodsky

A szenvedés a tudatosság egyedüli eredete.

-Fyodor Dosztojevszkij, *Feljegyzések a föld alól*

Van egy növényfajta, amely szerves táplálékot fogyaszt a virágaival: amikor egy légy telepszik a virágra, a szirmok rácsukódnak, és addig tartják, amíg a növény fel nem szívja a rovar a szervezetébe; de csak arra csukódnak rá, ami jó, amit meg lehet enni; egy esőcsepről vagy egy darab botról nem vesznek tudomást. Különös, hogy egy ilyen öntudatlan lény ilyen éles szemmel figyeli a saját érdekeit. Ha ez az öntudatlanság, akkor mi haszna van a tudatosságnak?

-Samuel Butler, 1871

Az agyat mint olyan entitást vizsgáltuk, amely bizonyos szintű teljesítményekre képes. De ez a nézőpont lényegében kihagyja *önmagunkat* a képből. Úgy tűnik, hogy az agyunkban élünk. Szubjektív életünk van. Hogyan viszonyul az agy objektív szemlélete, amelyet eddig tárgyaltunk, a saját érzéseinkhez, ahhoz az érzésünkhöz, hogy mi vagyunk az a személy, aki az élményeket átéli?

Colin McGinn brit filozófus (született 1950-ben) azt írja, hogy a "tudatosságról való vita még a legigényesebb gondolkodót is összefüggéstelenségbe taszíthatja". Ennek oka az, hogy az embereknek gyakran nincsenek átvilágított és ellentmondásos nézeteik arról, hogy pontosan mit is jelent a fogalom.

Sok megfigyelő a tudatot egyfajta teljesítménynek tekinti - például az önreflexió képességét, vagyis azt a képességet, hogy megértsük saját gondolatainkat és megmagyarázzuk azokat. Én ezt úgy írnám le, mint a saját gondolkodásunkról való gondolkodás képességét. Feltehetően ki tudnánk találni egy módszert ennek a képességnek az értékelésére, és aztán ezt a tesztet felhasználhatnánk arra, hogy elkülönítsük a tudatos dolgokat a tudatlan dolgoktól.

Azonban gyorsan bajba kerülünk, ha megpróbáljuk ezt a megközelítést megvalósítani. Tudatos a baba? Egy kutya? Nem nagyon tudják leírni a saját gondolkodási folyamatukat. Vannak emberek, akik úgy vélik, hogy a csecsemők és a kutyák éppen azért nem tudatos lények, mert nem tudják megmagyarázni magukat. Mi a helyzet a Watson néven ismert számítógéppel? Olyan üzemmódba kapcsolható, amelyben valóban megmagyarázza, hogyan jutott egy adott válaszra. Mivel saját gondolkodásának modelljét tartalmazza, Watson tehát tudatos, míg a csecsemő és a kutya nem az?

Mielőtt tovább elemeznénk ezt a kérdést, fontos elgondolkodnunk a kérdéssel kapcsolatos legjelentősebb megkülönböztetésen: Mi az, amit a tudományból megállapíthatunk, szemben azzal, ami valóban filozófia kérdése marad? Az egyik nézet szerint a filozófia egyfajta félmegoldás azoknak a kérdéseknek, amelyek még nem engedtek a tudományos módszernek. Eszerint, ha a tudomány eléggé előrehaladt ahhoz, hogy egy adott kérdéscsoportot megoldjon, a filozófusok továbbléphetnek más kérdések felé, amíg a tudomány meg nem oldja azokat is. Ez a nézet a tudatosság kérdésében, és különösen a "Mi és ki a tudatos?" kérdésben végigvonul.

Gondoljunk csak John Searle filozófus alábbi kijelentéseire: "Tudjuk, hogy az agy sajátos biológiai mechanizmusok segítségével tudatosságot okoznak. A lényeg az, hogy felismerjük, hogy a tudat olyan biológiai folyamat, mint az emésztés, a szoptatás, a fotoszintézis vagy a mitózis..... Az agy egy gép, egy biológiai gép, hogy biztos legyek benne, de attól még egy gép. Az első lépés tehát az, hogy kitaláljuk, hogyan csinálja ezt az agy, majd építünk egy olyan mesterséges gépet, amely ugyanilyen hatékony mechanizmussal rendelkezik a tudat előidézésére."¹ Az emberek gyakran meglepődnek, amikor ezeket az idézeteket látják, mert azt feltételezik, hogy Searle elkötelezett híve a tudat misztériumának védelmében az olyan redukcionistaikkal szemben, mint Ray Kurzweil.

David Chalmers ausztrál filozófus (született 1966-ban) a "tudat nehéz problémája" kifejezést használta, hogy leírja, milyen nehéz ezt az alapvetően leírhatatlan fogalmat meghatározni. Néha egy rövid mondat olyan jól összefoglal egy egész gondolkodási irányzatot, hogy emblematikussá válik (például Hannah Arendt "a gonosz banalitása"). Chalmers híres megfogalmazása ezt nagyon jól megvalósítja.

Amikor a tudatosságról beszélünk, nagyon könnyen belecsúszunk abba, hogy a megfigyelhető és mérhető tulajdonságokat vesszük figyelembe, amelyeket a tudatossághoz társítunk, de ez a megközelítés elhibázza az elképzelés lényegét. Az imént említettem a metakogníció fogalmát - a saját gondolkodásról való gondolkodás gondolatát -, amely a tudatosság egyik ilyen korrelátuma. Más megfigyelők az érzelmi intelligenciát vagy az

erkölcsi intelligenciát összemossák a tudattal. De ismétlem, az a képességünk, hogy kifejezzünk egy szeretetteljes érzést, hogy értjük a viccet, vagy hogy szexisek legyünk, egyszerűen teljesítménytípusok - talán lenyűgöző és intelligens, de olyan képességek, amelyek mindazonáltal megfigyelhetők és mérhetők (még ha vitatkozunk is arról, hogyan értékeljük őket). Annak kiderítése, hogy az agy hogyan hajtja végre az ilyen típusú feladatokat, és mi történik az agyban, amikor ezeket végezzük, képezi Chalmers "könnyű" tudati kérdését. Természetesen a "könnyű" probléma minden, csak nem az, és korunk talán legnehezebb és legfontosabb tudományos feladatát jelenti. Chalmers "nehéz" kérdése eközben annyira nehéz, hogy lényegében kimondhatatlan.

E különbségtétel alátámasztására Chalmers egy gondolat kísérletet mutat be, amely az általa zombiknak nevezett dolgokat érinti. A zombi egy olyan entitás, amely ugyanúgy viselkedik, mint egy ember, de egyszerűen nincs szubjektív tapasztalata - vagyis a zombi nem tudatos. Chalmers azzal érvel, hogy mivel el tudjuk képzelni a zombikat, legalábbis logikailag lehetségesek. Ha egy koktélpartin lennél, és ott lennének "normális" emberek és zombik is, hogyan tudnád megkülönböztetni őket? Talán ez úgy hangzik, mint egy koktélparti, amelyen ön is részt vett.

Sokan úgy válaszolnak erre a kérdésre, hogy az általuk értékelni kívánt személyeket az eseményekre és eszmékre adott érzelmi reakcióikról kérdeznék ki. Úgy vélik, hogy egy zombi bizonyos típusú érzelmi reakciók hiányával árulná el a szubjektív tapasztalat hiányát. Az ilyen irányú válasz azonban egyszerűen nem értékeli a gondolat kísérlet feltételezéseit. Ha egy érzelemmentes személlyel (például egy bizonyos érzelmi hiányosságokkal rendelkező egyénnel, mint az autizmus bizonyos típusainál gyakori) vagy egy avatárral vagy egy robottal találkozoznánk, amely nem lenne meggyőző, mint egy érzelmes emberi lény, akkor ez az entitás nem zombi. Ne feledje! Chalmers feltételezése szerint egy zombi teljesen normális reakcióképességgel rendelkezik, beleértve az érzelmi reakcióképességet is; csak hiányzik belőle a szubjektív tapasztalat. A lényeg az, hogy nincs mód a zombi azonosítására, mert a definíció szerint a viselkedésében nincs nyilvánvaló jele a zombi jellegének. Tehát ez egy különbségtétel különbség nélkül?

Chalmers nem próbálja megválaszolni a nehéz kérdést, de ad néhány lehetőséget. Az egyik a dualizmus egy formája, amelyben a tudat önmagában nem a fizikai világban, hanem inkább egy különálló ontológiai valóságként létezik. E megfogalmazás szerint az, amit egy személy tesz, az agyában zajló folyamatokon alapul. Mivel az agy kauzálisan zárt, egy személy cselekedeteit, beleértve a gondolatait is, teljes mértékben meg tudjuk magyarázni a folyamatai révén. A tudat tehát lényegében egy másik birodalomban létezik, vagy legalábbis a fizikai világtól elkülönült tulajdonság. Ez a magyarázat nem teszi lehetővé, hogy az elme (vagyis az agyhoz kapcsolódó tudati tulajdonság) kauzálisan befolyásolja az agyat.

A Chalmers által felvetett másik lehetőség, amely logikailag nem különbözik a dualizmusról alkotott elképzelésétől, és amelyet gyakran pánprotopszichizmusnak neveznek, azt állítja, hogy minden fizikai rendszer tudatos, bár az ember tudatosabb, mint mondjuk egy villanykapcsoló. Azzal mindenképpen egyetértenék, hogy az emberi agynak több tudatos dolga van, mint egy villanykapcsolónak.

Az én nézetem, amely talán a pánprotopszichizmus egyik aliskolája, az, hogy a tudatosság egy komplex fizikai rendszer emergens tulajdonsága. Eszerint a kutya is tudatos, de valamivel kevésbé, mint az ember. Egy hangya is rendelkezik valamilyen szintű tudattal,

de sokkal kevésbé, mint egy kutya. A hangyakolónia viszont magasabb szintű tudattal rendelkezik, mint az egyes hangyák; minden bizonnyal intelligensebb, mint egy magányos hangya. Eszerint a számítás szerint egy olyan számítógép, amely sikeresen utánozza az emberi agy komplexitását, ugyanolyan kialakuló tudatossággal rendelkezne, mint az ember.

A tudatosság fogalmát egy másik módon úgy is felfoghatjuk, mint egy olyan rendszert, amely "kvalitásokkal" rendelkezik. Mik is azok a kvaliák? A kifejezés egyik meghatározása a "tudatos élmények". Ez azonban nem visz bennünket túl messzire. Gondoljunk csak a következő gondolat kísérletre: Egy idegtudós teljesen színvak - nem az a fajta színvaktság, amikor az ember összekeveri mondjuk a zöld és a piros bizonyos árnyalatait (mint én), hanem inkább egy olyan állapot, amikor az érintett egyén teljesen fekete-fehér világban él. (Ennek a foratókönyvnek egy szélsőségesebb változata, hogy az illető fekete-fehér világban nőtt fel, és soha nem látott színeket. Lényegében nincs szín az ő világában). Ugyanakkor alaposan tanulmányozta a színek fizikáját - tisztában van vele, hogy a vörös fény hullámhossza 700 nanométer -, valamint a színeket normálisan megtapasztalni képes ember neurológiai folyamatait, és így sokat tud arról, hogyan dolgozza fel az agy a színeket. Többet tud a színekről, mint a legtöbb ember. Ha segíteni akarnál neki, és el akarnád magyarázni, hogy milyen is ez a tényleges "vörös" élmény, hogyan tennéd?

Talán felolvashatna neki egy részletet a nigériai költő, Oluseyi Oluseun "Vörös" című verséből:

*Vörös a vér színe az élet
szimbóluma*

*A vörös a veszély színe a halál
szimbóluma*

*Piros a rózsza színe a szépség
szimbóluma Piros a szerelmesek
színe az egység szimbóluma*

*Piros a paradicsom színe a jó
egészség szimbóluma Piros a forró
tűz színe
az égő vágy szimbóluma*

Ez tulajdonképpen elég jó képet adna neki arról, hogy az emberek milyen asszociációkat keltenek a vörös színnel kapcsolatban, és talán még azt is lehetővé tenné számára, hogy megállja a helyét egy a színről szóló beszélgetésben ("Igen, szeretem a vörös színt, olyan forró és tüzes, olyan veszélyesen szép...") Ha ő

ha akarta volna, valószínűleg meg tudta volna győzni az embereket arról, hogy vöröset tapasztalt, de a világ összes költészete nem tette volna lehetővé, hogy valóban átélje ezt az élményt.

Hasonlóképpen, hogyan magyarázná el, milyen érzés vízbe merülni valakinek, aki még soha nem érintette a vizet? Ismét kénytelenek lennénk a költészethez folyamodni, de magát az élményt nem igazán lehet átadni. Ezeket az élményeket nevezzük kvaliáknak.

E könyv olvasói közül sokan megtapasztalták már a piros színt. De honnan tudhatom, hogy a vörös színnel kapcsolatos élményük nem ugyanaz-e, mint amit én tapasztalok, amikor a kéket nézem? Mindketten ránézünk egy piros tárgyra, és biztosan állítjuk, hogy az piros, de ez még nem válaszolja meg a kérdést. Lehet, hogy én azt tapasztalom, amit te tapasztalsz, amikor a kékre nézel, de mindketten megtanultuk, hogy a piros dolgokat pirosnak nevezzük. Megint elkezdhetnénk verseket cserélgetni, de ezek csak az emberek színekkel kapcsolatos asszociációit tükröznék; nem a kvalitások tényleges természetéről szólnak. Valóban, a veleszületetten vak emberek sokat olvastak a színekről, mivel az irodalomban bőven találunk ilyen utalásokat, és így valóban rendelkeznek a színekkel kapcsolatos tapasztalat valamilyen változatával. Hogyan viszonyul a vörösről szerzett tapasztalatuk a látó emberek tapasztalatához? Ez tulajdonképpen ugyanaz a kérdés, mint a fekete-fehér világban élő nőre vonatkozó kérdés. Figyelemre méltó, hogy életünk ilyen gyakori jelenségei annyira teljesen kifejezhetetlenek, hogy lehetetlenné teszik az egyszerű megerősítést, például azt, hogy ugyanazokat a kvaliákat tapasztaljuk.

A qualia egy másik definíciója az élmény érzése. Ez a meghatározás azonban nem kevésbé körkörös, mint a tudat fenti meghatározási kísérleteink, mivel az "ézés", az "élmény" és a "tudat" kifejezések mind szinonimák. A tudat és a vele szorosan összefüggő qualia kérdése alapvető, talán a végső filozófiai kérdés (bár az identitás kérdése talán még fontosabb, ahogyan azt e fejezet záró részében tárgyalni fogom).

Tehát a tudatossággal kapcsolatban pontosan mi is a kérdés? Ez a következő: Ki vagy mi a tudatos? E könyv címében azért utalok az "elme" kifejezésre az "agy" helyett, mert az elme egy olyan agy, amely tudatos. Azt is mondhatnánk, hogy az elme szabad akarattal és identitással rendelkezik. Az az állítás, hogy ezek a kérdések filozófiaiak, önmagában nem magától értetődő. Állítom, hogy ezeket a kérdéseket a tudomány soha nem tudja teljes mértékben megoldani. Más szavakkal, nincsenek olyan hamisítható kísérletek, amelyeket mérlegelni tudnánk, amelyek megoldanák őket, filozófiai feltételezések nélkül nem. Ha tudatdetektort építenénk, Searle azt akarná, hogy meggyőződjön arról, hogy biológiai neurotranszmittereket spriccel. Daniel Dennett amerikai filozófus (szül. 1942) rugalmasabban viszonyulna a szubsztrátumhoz, de talán azt szeretné megállapítani, hogy a rendszer tartalmaz-e modellt önmagáról és saját teljesítményéről. Ez a nézet közelebb áll az enyémhez, de a lényege még mindig egy filozófiai feltételezés.

Rendszeresen kerülnek elő olyan javaslatok, amelyek olyan tudományos elméleteknek tűnnek, amelyek a tudatot valamilyen mérhető fizikai tulajdonsághoz kötik - amit Searle a "tudatosságot kiváltó mechanizmusnak" nevez. Stuart Hameroff (1947-ben született) amerikai tudós, filozófus és aneszteziológus azt írta, hogy

"A citoszkeletális filamentumok a tudat gyökerei."² A minden sejtben (beleértve a

neuronokat is, de nem csak azokban) található vékony szálakra, az úgynevezett mikrotubulusokra utal, amelyek minden sejtnak szerkezeti integritást adnak, és szerepet játszanak a sejtosztódásban. Az erről szóló könyvei és tanulmányai részletes leírásokat és egyenleteket tartalmaznak, amelyek megmagyarázzák annak a valószínűségét, hogy a mikrotubulusok szerepet játszanak a sejten belüli információfeldolgozásban. A mikrotubulusok és a tudatosság összekapcsolása azonban hitbeli ugrást igényel, amely nem különbözik alapvetően attól a hitbeli ugrástól, amely egy olyan vallási tanításban rejlik, amely egy felsőbb lényt ír le, amely tudatosságot (néha "léleknek" nevezik) adományoz bizonyos (általában emberi) entitásoknak. Hameroff nézete mellett néhány gyenge bizonyítékot is felhoztak, nevezetesen azt a megfigyelést, hogy az idegrendszeri folyamatok, amelyek támogathatják ezt az állítólagos sejtes számítástechnikát, leállnak az altatás során. Ez azonban messze nem meggyőző bizonyíték, tekintve, hogy rengeteg folyamat áll le az altatás során. Még azt sem állíthatjuk biztosan, hogy az alanyok nem tudatosak, amikor elaltatják őket. Annyit tudunk, hogy az emberek nem emlékeznek az élményeikre utána. Még ez sem általános, mivel néhány ember emlékszik - pontosan - az altatás alatt szerzett élményeire, beleértve például a sebészükkel folytatott beszélgetéseket is. Ezt a jelenséget aneszteziológiai tudatosságnak nevezik, és becslések szerint évente körülbelül 40 000 alkalommal fordul elő az Egyesült Államokban.³ De még ha ezt félretesszük is, a tudatosság és az emlékezet teljesen különböző fogalmak. Amint azt már részletesen megvitattam, ha visszagondolok az elmúlt nap során szerzett pillanatnyi élményeimre, rengeteg érzékszervi benyomásom volt, mégis nagyon kevésre emlékszem közülük. Nem voltam tehát tudatában annak, amit egész nap láttam és hallottam? Ez tulajdonképpen egy jó kérdés, és a válasz nem is olyan egyértelmű.

Roger Penrose angol fizikus és matematikus (született 1931-ben) másfajta hitet tett, amikor a tudat forrását javasolta, bár az ő javaslata szintén a mikrotubulusokra vonatkozott - pontosabban azok állítólagos kvantumszámítási képességeire. Érvelése, bár nem mondta ki kifejezetten, úgy tűnt, hogy a tudat titokzatos, és egy kvantumeseemény is titokzatos, tehát valamilyen módon össze kell kapcsolódnuk.

Penrose Turing megoldhatatlan problémákra vonatkozó tételeivel és Gödel ezzel kapcsolatos befejezetlenségi tételével kezdte elemzését. Turing előfeltevése (amelyet a [8](#) fejezetben részletesebben tárgyaltunk) az, hogy vannak olyan algoritmikus problémák, amelyeket ki lehet mondani, de amelyeket egy Turing-gép nem tud megoldani. A Turing-gép számítási univerzalitását tekintve arra következtethetünk, hogy ezeket a "megoldhatatlan problémákat" semmilyen gép nem tudja megoldani. Gödel befejezetlenségi tételének hasonló eredménye van a számokkal kapcsolatos feltevések bizonyítási képességére vonatkozóan. Penrose érvelése szerint az emberi agy képes megoldani ezeket a megoldhatatlan problémákat, tehát olyan dolgokra is képes, amelyekre egy determinisztikus gép, például egy számítógép nem képes. Motivációja, legalábbis részben, az, hogy az embert a gépek fölé emelje. De központi előfeltevése - miszerint az ember képes megoldani Turing és Gödel megoldhatatlan problémáit - sajnos egyszerűen nem igaz.

Egy híres megoldhatatlan probléma, az elfoglalt hód probléma a következőképpen hangzik: Egy bizonyos számú állapottal rendelkező Turing-gép maximálisan hány 1-est írhat a szalagra. Tehát az n szám busy beaverjének meghatározásához megalkotjuk az összes

olyan Turing-gépet, amelynek n állapota van (ami véges szám lesz, ha n véges), majd meghatározzuk a legnagyobb számú 1-est, amit ezek a gépek a szalagjukra írnak, kizárva azokat a Turing-gépeket, amelyek végtelen ciklusba kerülnek. Ez azért megoldhatatlan, mert mivel az összes ilyen n állapotú Turing-gépet próbáljuk szimulálni, a szimulátorunk végtelen ciklusba kerül, amikor megpróbálja szimulálni az egyik olyan Turing-gépet, amelyik végtelen ciklusba kerül. Kiderül azonban, hogy a számítógépek ennek ellenére képesek voltak meghatározni az elfoglalt hódfüggvényt bizonyos n -ekre. Az emberek is, de a számítógépek sokkal több n -re oldották meg a problémát, mint a segítség nélküli emberek. A számítógépek általában jobbak az embereknél a Turing és Gödel megoldhatatlan problémáinak megoldásában.

Penrose az emberi agy ezen állítólagos transzcendens képességeit az általa feltételezett kvantumszámítással hozta összefüggésbe. Penrose szerint ezek az idegi kvantumhatások valahogy eredendően nem voltak elérhetők a számítógépek által, ezért az emberi gondolkodásnak eredendően előnye van. Valójában a közönséges elektronika kvantumhatásokat használ (a tranzisztorok az elektronok kvantum-alagutazására támaszkodnak az akadályokon keresztül); az agyban történő kvantumszámítást soha nem sikerült bizonyítani; az emberi szellemi teljesítményt a klasszikus számítási módszerekkel kielégítően meg lehet magyarázni; és mindenesetre semmi sem zárja ki, hogy a számítógépekben kvantumszámítást alkalmazzunk. Penrose ezen ellenvetések egyikével sem foglalkozott. Hameroff és Penrose akkor fogott össze, amikor a kritikusok rámutattak, hogy az agy meleg és kusza hely a kvantumszámítás számára. Penrose megtalálta a neuronokon belül a tökéletes hordozót, amely elképzelhetően támogathatja a kvantumszámítást - nevezetesen a mikrotubulusokat, amelyekről Hameroff azt feltételezte, hogy a neuronon belüli információfeldolgozás részei. A Hameroff-Penrose tézis szerint tehát a neuronokban lévő mikrotubulusok végzik a kvantumszámítást, és ez felelős a tudatosságért.

Ezt a tézist kritizálta például Max Tegmark (1967-ben született) svéd amerikai fizikus és kozmológus is, aki megállapította, hogy a mikrotubulusokban zajló kvantumesemények csak 10^{-13} másodpercig maradhatnak fenn, ami túl rövid idő ahhoz, hogy bármilyen jelentőségű eredményeket számítsanak ki, vagy hogy befolyásolják az idegrendszer működését.

folyamatok. Vannak bizonyos problématípusok, amelyek esetében a kvantumszámítástechnika a klasszikus számítástechnikához képest jobb képességekkel rendelkezik.

-például a titkosítási kódok feltörése nagy számok faktorálásával. A segítség nélküli emberi gondolkodás azonban szörnyűnek bizonyult ezek megoldásában, és ezen a területen még a klasszikus számítógépekkel sem képes felvenni a versenyt, ami arra utal, hogy az agy nem mutat kvantumszámítási képességeket. Sőt, még ha létezne is az agyban olyan jelenség, mint a kvantumszámítás, az nem feltétlenül kapcsolódna a tudatossághoz.

Hinned kell a hitben

Micsoda munka ez az ember! Milyen nemes észjárású! Milyen végtelen képességekkel! Milyen kifejező és csodálatra méltó a formája és a mozgása! Tetteiben mennyire hasonlít egy angyalra! Felfogásában mennyire hasonlít egy istenhez! A világ szépsége! Az állatok példaképe! És mégis, számomra mi ez a pornak ez a kvintesszenciája?

-Hamlet, Shakespeare *Hamletjében*

A valóság az, hogy ezek az elméletek mind a hit ugrásai, és hozzátenném, hogy a tudatossággal kapcsolatban a vezérelv az, hogy "hinned kell" - vagyis mindannyiunknak szüksége van egy hitugrásra, hogy mi és ki a tudatos, és kik és mik vagyunk mi, mint tudatos lények. Máskülönben nem tudnánk reggel felkelni. De őszintének kell lennünk azzal kapcsolatban, hogy ebben a kérdésben alapvetően szükségünk van a hit ugrására, és önmagunknak kell lennünk azzal kapcsolatban, hogy a saját konkrét ugrónk mit foglal magában.

Az emberek nagyon különbözően ugranak, az ellenkező benyomások ellenére. A tudat természetével és forrásával kapcsolatos egyéni filozófiai feltevések állnak az állatok jogaitól az abortuszig terjedő kérdésekben fennálló nézeteltérések háttérben, és a jövőben még vitatottabb konfliktusokhoz fognak vezetni a gépek jogai miatt. Objektív jóslatom az, hogy a gépek a jövőben tudatosnak fognak tűnni, és meggyőzőek lesznek a biológiai emberek számára, amikor a kvalitásaikról beszélnek. A finom, ismerős érzelmi jelzések teljes skáláját mutatják majd; megnevetetnek és megríkatnak minket; és meg fognak haragudni ránk, ha azt mondjuk, hogy nem hisszük el, hogy tudatosak. (Nagyon okosak lesznek, ezért nem akarjuk majd, hogy ez megtörténjen.) El fogjuk fogadni, hogy tudatos személyek. Az én saját hitem a következő: Ha egyszer a gépeknek sikerül meggyőzőnek lenniük, amikor a kvalitásaikról és a tudatos tapasztalataikról beszélnek, akkor valóban tudatos személyeknek fognak minősülni. Erre az álláspontomra ezen a gondolatkísérleten keresztül jutottam el: Képzeld el, hogy a jövőben találkozunk egy olyan entitással (egy robottal vagy egy avatárral), amely teljesen meggyőző az érzelmi reakcióiban. Meggyőzően nevet a vicceiden, és cserébe megnevetet és megríkat téged (de nem csak azzal, hogy megcsíp). Meggyőző az őszinteségéről, amikor a félelmeiről és vágyairól beszél. Minden tekintetben tudatosnak tűnik. Valójában úgy tűnik, mintha ember lenne. Elfogadnád őt tudatos személynek?

Ha a kezdeti reakció az, hogy valószínűleg észreveszel valamilyen módon, amiben elárulja a nem biológiai természetét, akkor nem tartod be a feltételezéseket ebben a hipotetikus helyzetben, amely megállapította, hogy a nő teljesen meggyőző. E feltételezés alapján, ha őt pusztulással fenyegetnék, és erre, ahogyan egy ember tenné, rémülettel reagálna, ugyanolyan empatikusan reagálnál, mintha egy ilyen jelenetnek lennél szemtanúja, amelyben egy ember is érintett? Számomra a válasz igen, és azt hiszem, a válasz ugyanez lenne a legtöbb, ha nem gyakorlatilag minden más ember számára is, függetlenül attól, hogy mit állítanak most egy filozófiai vitában. Ismétlem, a hangsúly itt a "meggyőző" szón van.

Bizonyára nem egységesesek a vélemények arról, hogy mikor találkoznak egy ilyen nem biológiai entitással, vagy egyáltalán találkoznak-e vele. Az én következetes előrejelzésem szerint ez először 2029-ben fog bekövetkezni, és a 2030-as években válik rutinná. De félretéve az időkeretet, úgy vélem, hogy előbb-utóbb el fogunk jutni oda, hogy az ilyen entitásokat tudatosnak tekintjük. Gondoljunk csak arra, hogyan bánunk velük már most is, amikor mesék és filmek szereplőiként találkoznak velük: R2D2 a *Star Wars*-filmekből, David és Teddy az *A.I.* című filmből, Data a *Star Trek: The Next Generation* című sorozatból, Johnny 5 a *Short Circuit* című filmből, WALL-E a Disney *Wall-E* című filmjéből, T-800 - a (jó) Terminátor - a második és a későbbi Terminátor-filmekből, Rachael a Replikáns a *Pengefutó* című filmből (aki egyébként nem tudja, hogy nem ember), Bumblebee a *Transformers* című film-, tévé- és képregénysorozatból, és Sonny az *I, Robot* című filmből. Együtt érzünk ezekkel a karakterekkel, még akkor is, ha tudjuk, hogy nem biológiailag léteznek. Ugyanúgy tudatos személyeknek tekintjük őket, mint a biológiai emberi karaktereket. Osztozunk az érzéseikben, és féltjük őket, amikor bajba kerülnek. Ha ma így bánunk a fiktív, nem biológiai karakterekkel, akkor a jövőben így fogunk bánni a valós életben élő intelligenciákkal, amelyeknek történetesen nincs biológiai szubsztrátjuk.

Ha elfogadod azt a hitbeli ugrást, hogy egy nem biológiai entitás, amely meggyőzően reagál a kvalitásokra, valóban tudatos, akkor fontold meg, hogy ez mit jelent: nevezetesen azt, hogy a tudatosság egy entitás általános mintázatának emergens tulajdonsága, nem pedig a szubsztrátumé, amelyen fut.

Van egy fogalmi szakadék a tudomány, amely az *objektív* mérést és az ebből levonható következtetéseket jelenti, és a tudatosság, amely a *szubjektív* tapasztalat szinonimája. Nyilvánvalóan nem tudjuk egyszerűen megkérdezni egy kérdéses entitástól, hogy "Ön tudatánál van-e?". Ha belenézünk a "fejébe", legyen az biológiai vagy más, hogy ezt megállapítsuk, akkor filozófiai feltevéseket kellene tennünk annak meghatározásához, hogy mi az, amit keresünk. A kérdés, hogy egy entitás tudatos-e vagy sem, tehát nem tudományos kérdés. Ebből kiindulva egyes megfigyelők továbbmennek, és megkérdőjelezik, hogy magának a tudatnak van-e bármilyen valóságalapja. Susan Blackmore angol író és filozófus (szül. 1951) a "tudatosság nagy illúziójáról" beszél. Elismeri a tudat mémjének (eszméjének) valóságát - más szóval, a tudat mint eszme bizonyosan létezik, és rengeteg neokortikális struktúra foglalkozik az eszmével, nem is beszélve a róla elhangzott és leírt szavakról. De nem egyértelmű, hogy valami valóságos dologra utal. Blackburn a továbbiakban kifejti, hogy nem feltétlenül tagadja a tudat valóságát, hanem inkább megpróbálja megfogalmazni azokat a fajta dilemmákat, amelyekkel akkor találkoznak, amikor megpróbáljuk behatárolni a fogalmat. Ahogy Stuart Sutherland (1927-1998) brit pszichológus és író írta az *International Dictionary of Psychology* című könyvében: "A tudat lenyűgöző, de megfoghatatlan jelenség; lehetetlen meghatározni, hogy mi az, mit csinál, vagy miért alakult ki".⁴

Azonban jól tesszük, ha nem utasítjuk el a fogalmat, hogy egyszerűen csak filozófusok közötti udvarias vitaként kezeljük - ami egyébként kétezer évvel ezelőttre, a platóni párbeszédekre nyúlik vissza. A tudatosság eszméje az erkölcsi rendszerünk alapja, és a jogrendszerünk viszón lazán ezekre az erkölcsi meggyőződésekre épül. Ha valaki kioltja valakinek a tudatát, mint például a gyilkoságnál, azt erkölcsstelennek tartjuk, és néhány kivételtől eltekintve súlyos bűncselekménynek. Ezek a kivételek a tudatosságra is

vonatköznak, amennyiben engedélyezhetjük, hogy a rendőrség vagy a hadsereg bizonyos tudatos embereket megöljön, hogy nagyobb számú más tudatos embert megvédjen. Vitatkozhatunk az egyes kivételek előnyeiről, de az alapelv igaz.

Az, hogy valakit bántalmaznak és szenvedést okoznak neki, szintén általában erkölcstelennek és törvénytelennek számít. Ha tönkreteszem a tulajdonomat, az valószínűleg elfogadható. Ha az engedélyed nélkül pusztítom el a tulajdonodat, az valószínűleg nem elfogadható, de nem azért, mert szenvedést okozok a tulajdonodnak, hanem inkább neked, mint a tulajdon tulajdonosának. Másrészt, ha a tulajdonomban egy tudatos lény, például egy állat van, akkor nekem, mint az állat tulajdonosának nem feltétlenül van szabad erkölcsi vagy jogi jogom arra, hogy azt tegyek vele, amit akarok - vannak például törvények az állatkínzás ellen.

Mivel erkölcsi és jogi rendszerünk nagy része azon alapul, hogy megvédjük a tudatos entitások létezését és megakadályozzuk a tudatos entitások szükségtelen szenvedését, a felelős ítélkezés érdekében választ kell adnunk arra a kérdésre, hogy ki a tudatos. Ez a kérdés tehát nem egyszerűen intellektuális vita tárgya, ahogyan azt az olyan kérdéseket, mint az abortusz körüli vita is mutatja. Rá kell mutatnom, hogy az abortusz kérdése némileg túlmutathat a tudatosság kérdésén, mivel az életvédelem hívei azzal érvelnek, hogy az embrió potenciális lehetősége, hogy végül tudatos személyiséggé váljon, elegendő ok arra, hogy védelmet kapjon, ahogyan a kómában fekvő ember is megérdemli ezt a jogot. De alapvetően a kérdés arról szól, hogy a magzat mikor válik öntudatra.

A tudatosságról alkotott elképzelések gyakran befolyásolják ítéleteinket ellentmondásos területeken is. Ha ismét az abortusz kérdését nézzük, sokan különbséget tesznek egy olyan intézkedés, mint a reggeli utáni tabletta, amely megakadályozza az embrió beágyazódását a méhbe a terhesség első napjaiban, és a késői abortusz között. A különbség annak a valószínűségével kapcsolatos, hogy a késői stádiumban lévő magzat tudatánál van-e. Nehéz azt állítani, hogy egy néhány napos embrió tudatos, hacsak nem foglalunk el pánprotopszichista álláspontot, de még ilyen szempontból is a legegyszerűbb állatnál is alacsonyabb rendű lenne a tudatosság szempontjából. Hasonlóképpen, az emberszabású majmok bántalmazására egészen másképpen reagálunk, mint mondjuk a rovarokéra. Ma senki sem aggódik túlságosan amiatt, hogy fájdalmat és szenvedést okozunk a számítógépes szoftverünknek (bár sokat kommentáljuk a szoftverek azon képességét, hogy szenvedést okozzanak nekünk), de amikor a jövőbeli szoftverek a biológiai emberek intellektuális, érzelmi és erkölcsi intelligenciájával rendelkeznek, ez valódi aggodalomra ad okot.

Ezért az én álláspontom az, hogy elfogadom, hogy az érzelmi reakcióikban teljesen meggyőző nem biológiai entitások tudatos személyek legyenek, és azt jószólam, hogy a társadalmi konszenzus is el fogja fogadni őket. Megjegyzendő, hogy ez a meghatározás túlmutat azokon az entitásokon, amelyek át tudnak menni a Turing-teszten, amihez az emberi nyelv elsajátítása szükséges. Ez utóbbiak kellően emberszerűek ahhoz, hogy én is ide soroljam őket, és úgy vélem, hogy a társadalom nagy része is így tesz, de olyan entitásokat is ide sorolok, amelyek emberszerű érzelmi reakciókat mutatnak, de nem képesek átmenni a Turing-teszten - például a kisgyermeket.

Ez megoldja a filozófiai kérdést, hogy ki a tudatos, legalábbis számomra és mások számára, akik elfogadják ezt a bizonyos hitbeli ugrást? A válasz: *nem egészen*. Eddig csak egy esetet érintettünk, mégpedig az emberhez hasonlóan viselkedő entitások esetét. Bár

olyan jövőbeli entitásokról beszélünk, amelyek agya az emberéhez hasonló, mégis egészen másfajta viselkedést mutatnak, tehát ez az álláspont még mindig emberközpontú. De mi a helyzet az intelligencia idegenebb formáival, amelyek nem emberhez hasonlóak? Elképzelhetünk olyan intelligenciákat, amelyek ugyanolyan bonyolultak, vagy talán sokkal összetettebbek és bonyolultabbak, mint az emberi agy, de teljesen más érzelmekkel és motivációkkal rendelkeznek. Hogyan döntsük el, hogy tudatosak-e vagy sem?

Kezdhetjük a biológiai világ olyan élőlényeivel, amelyek agya az emberéhez hasonló, mégis egészen másfajta viselkedést mutatnak. David Cockburn brit filozófus (1949-ben született) arról ír, hogy látott egy videót egy óriás tintahalról, amelyet megtámadtak (vagy legalábbis azt hitte, hogy megtámadták - Cockburn feltételezte, hogy talán a videokamerás embertől félt). A tintahal megremegett és meghúzódott, és Cockburn így ír erről: "Olyan módon reagált, amely azonnal és erőteljesen félelemmel töltött el. Ebben a jelenetben részben az volt a lenyűgöző, hogy egy olyan lény viselkedésében, amely fizikailag annyira különbözik az embertől, olyan érzelmeket lehetett látni, mint amilyenek az emberek voltak.

amely olyan egyértelműen és konkrétan a félelemről szólt."⁵ Arra a következtetésre jut, hogy az állat ezt az érzelmet érezte, és kifejti azt a meggyőződését, hogy a filmet néző legtöbb ember ugyanerre a következtetésre jutna. Ha elfogadjuk Cockburn leírását és következtetését, akkor az óriáskalmárokat fel kell vennünk a tudatos entítások listájára. Azonban ezzel sem jutottunk túl messzire, mert ez még mindig egy olyan érzelemre adott empatikus reakciónkon alapul, amelyet magunkban felismerünk. Ez még mindig egy énközpontú vagy emberközpontú perspektíva.

Ha kilépünk a biológiából, a nem biológiai intelligencia még változatosabb lesz, mint a biológiai világ intelligenciája. Például egyes entítások nem félhetnek a saját pusztulásuktól, és talán nincs szükségük az emberekben vagy bármely biológiai lényben megfigyelhető érzelmekre. Talán még így is át tudnának menni a Turing-teszten, de az is lehet, hogy meg sem akarnák próbálni.

Valójában ma már olyan robotokat építünk, amelyeknek nincs öfenntartási érzékük, hogy veszélyes környezetben hajtsanak végre küldetéseket. Ezek még nem elég intelligensek vagy összetettek ahhoz, hogy komolyan elgondolkodjunk az érzőképességükön, de el tudjuk képzelni a jövő ilyen robotjait, amelyek ugyanolyan összetettek, mint az emberek. Mi a helyzet velük?

Személy szerint azt mondanám, hogy ha egy ilyen eszköz viselkedésében egy összetett és értékes cél iránti elkötelezettséget látnék, és azt a képességet, hogy figyelemre méltó döntéseket és cselekvéseket hajtson végre a küldetése teljesítése érdekében, akkor le lennék nyűgözve, és valószínűleg feldúlt lennék, ha megsemmisülne. Ez most talán egy kicsit túlzásba viszi a fogalmat, mivel olyan viselkedésre reagálok, amely nem tartalmaz számos olyan érzelmet, amelyet az embereknél, sőt mindenféle biológiai élőlénynél univerzálisnak tartunk. De megint csak olyan tulajdonságokhoz próbálok kapcsolódni, amelyekkel magamban és más emberekben is tudok azonosulni. Az az elképzelés, hogy egy entitás teljes mértékben egy nemes célnak szenteli magát, és azt a saját jólétére való tekintet nélkül hajtja végre, vagy legalábbis megpróbálja megvalósítani, végül is nem teljesen idegen az emberi tapasztalatoktól. Ebben az esetben is egy olyan entitásról van szó, amely a biológiai embereket igyekszik megvédeni, vagy valamilyen módon a mi napirendünket

előmozdítani. mi akkor is van, ha ennek az entitásnak az emberi céloktól eltérő saját céljai vannak, és nem olyan küldetést hajt végre, amelyet mi a magunk szempontjából nemesnek ismernénk el? Akkor talán megpróbálnám megnézni, hogy más módon tudnék-e kapcsolódni vagy értékelni néhány képességét. Ha valóban nagyon intelligens, akkor valószínűleg jól ért a matematikához, így talán beszélgetést folytathatnék vele erről a témáról. Talán értékelné a matematikai vicceket.

De ha az entitásnak nem áll érdekében kommunikálni velem, és nincs elegendő hozzáférésem a cselekvéseihez és a döntéshozatalához ahhoz, hogy meghatódjak belső folyamatainak szépségétől, akkor ez azt jelenti, hogy nem tudatos? Arra kell következtetnem, hogy azok az entitások, amelyeknek nem sikerül meggyőzniük engem az érzelmi reakcióikról, vagy amelyek nem törődnek azzal, hogy megpróbálják, nem feltétlenül nem tudatosak. Nehéz lenne felismerni egy másik tudatos entitást anélkül, hogy valamilyen szintű empatikus kommunikációt létesítenék, de ez az ítélet inkább a saját korlátaimat tükrözi, mint a szóban forgó entitást. Ezért alázattal kell eljárunk. Elég nagy kihívás egy másik ember szubjektív helyzetébe beleélni magunkat, így a feladat sokkal nehezebb lesz olyan intelligenciák esetében, amelyek rendkívül különböznek a miénktől.

Mit tudatosítunk?

Ha a koponyán keresztül belenézhetnénk egy tudatosan gondolkodó ember agyába, és ha az optimális ingerlékenység helye világító lenne, akkor az agyfelszínen egy fényes foltot látnánk, amelynek fantasztikus, hullámzó határai állandóan változnak méretben és formában, és amelyet többé-kevésbé mély sötétség vesz körül, amely a félteke többi részét elborítja.

-Ivan Petrovics Pavlov, 1913⁶

Visszatérve az óriáskalmárhoz, felismerhetjük néhány nyilvánvaló érzelmét, de viselkedésének nagy része rejtély. Milyen érzés óriáskalmárnak lenni? Mit érez, amikor gerinctelen testét egy apró nyíláson átréseli? Még a szókincsünk sincs meg ahhoz, hogy erre a kérdésre válaszoljunk, tekintve, hogy még azokat az élményeket sem tudjuk leírni, amelyekben más emberekkel osztozunk, például a piros színt látjuk, vagy érezzük, ahogy a víz a testünkre fröccsen.

De nem kell egészen az óceán fenekéig mennünk ahhoz, hogy rejtélyeket találjunk a tudatos tapasztalatok természetében - elég, ha csak a sajátunkra gondolunk. Én például tudom, hogy tudatos vagyok. Feltételezem, hogy ön, az olvasó, szintén tudatos. (Ami azokat illeti, akik nem vették meg a könyvem, ebben nem vagyok olyan biztos.) De *miben* vagyok tudatos? Ön is felteheti magának ugyanezt a kérdést.

Próbálja ki ezt a gondolat kísérletet (amely azok számára is működik, akik autót vezetnek): Képzeld el, hogy az autópálya bal sávjában vezetünk. Most csukja be a szemét, ragadjon meg egy elképzelt kormánykereket, és tegye meg a mozdulatokat, hogy sávot

váltson a jobb oldali sávba. Oké, mielőtt folytatnád az olvasást, próbáld ki. Valószínűleg a következőt csináltad: Megfogtad a kormánykereket. Ellenőrizted, hogy a jobb oldali sáv szabad-e. Feltételezve, hogy a sáv szabad, rövid időre jobbra fordítottad a kormányt. Aztán újra kiegyenesítetted. A munkát elvégezte.

Még jó, hogy nem egy igazi autóban ültél, mert az autópálya összes sávján átzúztál, és egy fának ütköztél. Bár valószínűleg meg kellett volna említenem, hogy ezt nem szabadna valódi mozgó autóban kipróbálnod (de akkor feltételezem, hogy már elsajátítottad azt a szabályt, hogy csukott szemmel nem szabad vezetni), de itt nem is ez a fő probléma. Ha az imént leírt eljárást alkalmaztad - és szinte mindenki ezt teszi, amikor ezt a gondolat kísérletet végzi -, akkor rosszul csináltad. Ha jobbra fordítod a kormányt, majd kiegyenesíted, akkor az autó az eredeti irányhoz képest átlós irányba megy. Átmegy a sávot jobbra, ahogyan önök szándékozták, de a végtelenségig jobbra fog menni, amíg le nem zúg az útról. Amit tennie kellett, amikor az autója jobbra keresztezte a sávot, az az volt, hogy ezután balra fordította a kormányt, éppen annyira, amennyire jobbra fordította, *majd* újra egyenesbe hozta. Ezáltal az autó ismét egyenesen fog haladni az új sávban.

Gondoljon arra a tényre, hogy ha Ön rendszeresen vezet, akkor ezt a manővert már több ezer alkalommal végrehajtotta. Nem vagy tudatában, amikor ezt csinálod? Soha nem figyeltél arra, hogy valójában mit csinálsz, amikor sávot váltasz? Feltételezve, hogy nem egy kórházban olvassa ezt a könyvet, miközben egy sávváltásos balesetből lábadozik, egyértelműen elsajátította ezt a készséget. Mégsem vagy tudatában annak, hogy mit tettél, akárhányszor is hajtottad végre ezt a feladatot.

Amikor az emberek élményeikről mesélnek, azokat helyzetek és döntések sorozataként írják le. De mi nem így éljük meg a történetet. Eredeti tapasztalatunk magas szintű minták sorozataként jelenik meg, amelyek közül néhány érzéseket válthat ki. Ezeknek a mintáknak csak egy kis részhalmozára emlékszünk, ha egyáltalán emlékszünk. Még ha viszonylag pontosan el is mesélünk egy történetet, a konfabulációs képességünket használjuk arra, hogy a hiányzó részleteket kitöltsük, és a sorozatot koherens történetté alakítsuk. Emlékezetünk alapján nem lehetünk biztosak abban, hogy mi volt az eredeti tudatos élményünk, mégis az emlékezet az egyetlen hozzáférésünk ehhez az élményhez. A jelen pillanat, nos, múlandó, és gyorsan emlékké alakul, vagy, ami még gyakoribb, nem. Még ha egy élmény emlékké is alakul, akkor is, ahogy a PRTM jelzi, egy hatalmas hierarchiában más mintákból álló, magas szintű mintaként tárolódik. Mint már többször rámutattam, szinte minden élményünket (mint például bármelyik alkalommal, amikor sávot váltottunk) azonnal elfelejtjük. Tehát annak megállapítása, hogy mi alkotja a saját tudatos tapasztalatainkat, valójában nem elérhető.

Kelet az Kelet és Nyugat az Nyugat

Az agyak előtt nem volt szín vagy hang a világegyetemben, sem íz vagy illat, és valószínűleg kevés érzék, érzés vagy érzelem.

-Roger W. Sperry²

René Descartes besétál egy étterembe, és leül vacsorázni. A pincér odajön hozzá, és megkérdezi, kér-e előételt.

"Nem, köszönöm - mondja Descartes -, csak vacsorát szeretnék rendelni."

"Szeretné hallani a napi ajánlatainkat?" - kérdezi a pincér.

"Nem - mondja Descartes türelmetlenül.

"Szeretne egy italt vacsora előtt?" - kérdezi a pincér. Descartes megsértődik, hiszen ő antialkoholista. "Azt hiszem, nem!"

mondja felháborodottan, és POOF! eltűnik.

-Egy vicc, ahogy David Chalmers felidézte

Az általunk vizsgált kérdéseket kétféleképpen lehet szemlélni - a tudat és a valóság természetének nyugati és keleti nézőpontjait. A nyugati perspektívában egy fizikai világból indulunk ki, amely információmintákat fejleszt. Néhány milliárd évnyi evolúció után a világban lévő entitások eléggé fejlődtek ahhoz, hogy tudatos lényekké váljanak. A keleti szemlélet szerint a tudat az alapvető valóság; a fizikai világ csak a tudatos lények gondolatai révén jön létre. Más szóval a fizikai világ a tudatos lények gondolatai, amelyek megnyilvánulnak. Ezek természetesen leegyszerűsítései az összetett és sokféle filozófiának, de ezek a tudatról és a fizikai világhoz való viszonyáról szóló filozófiák fő polaritásai.

A tudatosság kérdésének kelet-nyugati megosztottsága a szubatomi fizika területén is kifejezésre jutott az ellentétes gondolkodásmódban. A kvantummechanikában a részecskék úgynevezett valószínűségi mezőkként léteznek. Bármilyen mérőeszköz által rajtuk végzett mérés a hullámfüggvény úgynevezett összeomlását okozza, ami azt jelenti, hogy a részecske hirtelen egy bizonyos helyet vesz fel. Egy népszerű nézet szerint egy ilyen mérés egy tudatos megfigyelő általi megfigyelésnek minősül, mert különben a mérés értelmetlen fogalom lenne. A részecske tehát csak akkor vesz fel egy adott helyet (és más tulajdonságokat, például sebességet), ha megfigyelik. Alapvetően a részecskék úgy gondolják, hogy ha senki sem veszi a fáradságot, hogy megnézzék őket, akkor nem kell eldönteniük, hogy hol vannak. Én ezt a kvantummechanika buddhista iskolájának nevezem, mert ebben a kvantummechanikában a részecskék lényegében nem léteznek, amíg egy tudatos személy meg nem figyeli őket.

A kvantummechanikának van egy másik értelmezése is, amely kerüli az ilyen antropomorf terminológiát. Ebben az elemzésben a részecskét reprezentáló mező nem valószínűségi mező, hanem csak egy függvény, amely különböző helyeken különböző értékeket vesz fel. A mező tehát alapvetően az, ami a részecske. Korlátok vannak arra nézve, hogy a mező értékei milyenek lehetnek a különböző helyeken, mert a részecskét reprezentáló teljes mező csak korlátozott mennyiségű információt képvisel. Innen ered a "kvantum" szó. A hullámfüggvény úgynevezett összeomlása - e nézet szerint - egyáltalán nem összeomlás. A hullámfüggvény valójában soha nem tűnik el. Csupán arról van szó, hogy egy mérőeszköz is részecskékből áll, amelyeknek mezői vannak, és a mérendő részecskemező és a mérőeszköz részecskemezőinek kölcsönhatása azt eredményezi, hogy a részecske egy adott helyen van. A mező azonban továbbra is jelen van. Ez a kvantummechanika nyugati értelmezése, bár érdekes megjegyezni, hogy a fizikusok körében

világszerte népszerűbb nézet az, amit én keleti értelmzésemnek neveztem.

Volt egy filozófus, akinek a munkássága átívelte ezt a kelet-nyugati szakadékot. Ludwig Wittgenstein (1889- 1951) osztrák brit gondolkodó a nyelv és a tudás filozófiáját tanulmányozta, és azon elmélkedett, hogy mi az, amit valóban tudhatunk. Az első világháború katonájaként elmélkedett ezen a témán, és jegyzeteket készített egyetlen, még életében megjelent könyvéhez, a *Tractatus Logico-Philosophicus*hoz. A mű szokatlan szerkezetű volt, és csak korábbi oktatója, Bertrand Russell brit matematikus és filozófus erőfeszítései révén talált kiadóra 1921-ben. A mű a logikai pozitivizmus néven ismert nagy filozófiai iskola bibliájává vált, amely a tudomány határait igyekezett meghatározni. A könyv és az azt körülvevő mozgalom nagy hatással volt Turingra, valamint a számításmélet és a nyelvészet kialakulására.

A *Tractatus Logico-Philosophicus* megelőlegezi azt a felismerést, hogy minden tudás eredendően hierarchikus. Maga a könyv egymásba ágyazott és számozott állításokba rendeződik. A könyv első négy állítása például a következő:

1 A világ minden, ami a helyzet.

1.1 A világ a tények összessége, nem a dolgoké.

1.11 A világot a tények határozzák meg, és az, hogy ezek mind tények.

1.12 A tények összessége határozza meg, hogy mi az eset, és azt is, hogy mi nem az eset.

A *Tractatus* egy másik jelentős kijelentése - amelyet Turing is visszhangzott - ez:

4.0031 Minden filozófia a nyelv kritikája.

Lényegében mind a *Tractatus Logico-Philosophicus*, mind a logikai pozitivizmus mozgalma azt állítja, hogy a fizikai valóság külön létezik a róla való érzékelésünktől, de erről a valóságról csak annyit tudhatunk, amit érzékszerveinkkel érzékelünk - amelyeket eszközeinkkel fokozhatunk -, és azokat a logikai következtetéseket, amelyeket ezekből az érzékszervi benyomásokból levonhatunk. Wittgenstein lényegében a tudomány módszereit és céljait próbálja leírni. A könyv utolsó megállapítása a 7. szám: "Amiről nem tudunk beszélni, arról hallgatva kell elmennünk". A korai Wittgenstein ennek megfelelően a tudatról való vitát körkörösnek és tautologikusnak, tehát időpocsékolásnak tartja.

A későbbi Wittgenstein azonban teljesen elutasította ezt a megközelítést, és minden filozófiai figyelmét olyan dolgokról való beszédre fordította, amelyekről korábban azt állította, hogy hallgatni kellene róluk. Ennek az átdolgozott gondolkodásmódnak az írásait 1953-ban, két évvel a halála után, a *Filozófiai vizsgálódások* című könyvben gyűjtötték össze és adták ki. A *Tractatusban* kritizálta korábbi gondolatait, körkörösnek és értelmetlennek ítélve azokat, és arra a nézetre jutott, hogy amiről ő azt tanácsolta, hogy ne beszéljünk, az valójában mindaz, amiről érdemes elgondolkodni. Ezek az írások nagy hatással voltak az egzisztencialistákra, így Wittgenstein lett a modern filozófia egyetlen olyan alakja, aki a filozófia két vezető és egymásnak ellentmondó gondolkodási irányzatának fő tervezője.

Mi az, amiről a későbbi Wittgenstein szerint érdemes volt gondolkodni és beszélni?

Olyan kérdésekről, mint a szépség és a szerelem, amelyekről felismerte, hogy tökéletlenül léteznek ideaként az emberek elméjében. Ugyanakkor azt írja, hogy az ilyen fogalmak valóban léteznek egy tökéletes és idealizált birodalomban, hasonlóan a tökéletes "formákhoz", amelyekről Platón írt a platóni dialógusokban, egy másik művében, amely a valóság természetének látszólag ellentmondásos megközelítését világította meg.

Az egyik gondolkodó, akinek álláspontját szerintem rosszul jellemzik, René Descartes francia filozófus és matematikus. Híres "Gondolkodom, tehát vagyok" mondatát általában úgy értelmezik, hogy a racionális gondolkodást dicsőíti, abban az értelemben, hogy "gondolkodom, azaz képes vagyok logikus gondolkodást végezni, tehát érdemes vagyok". Descartes-t ezért a nyugati racionális szemlélet megalkotójának tartják.

Ezt a kijelentést más írásai kontextusában olvasva azonban más benyomást kelt bennem. Descartes-t nyugtalanította az, amit "test és elme problémájának" neveznek: Nevezetesen, hogyan keletkezik a tudatos elme az agy fizikai anyagából? Ebből a szempontból úgy tűnik, hogy a racionális szkepticizmust próbálta a végsőkéig feszegetni, így véleményem szerint kijelentése valójában azt jelenti: "Azt hiszem, vagyis szubjektív élmény történik, ezért csak annyit tudunk biztosan, hogy valami - nevezzük *énnek* - létezik". Nem lehetett biztos abban, hogy a fizikai világ létezik, mert csak a saját egyéni érzéki benyomásaink vannak róla, amelyek tévesek vagy teljesen illuzórikusak lehetnek. Azt azonban tudjuk, hogy a tapasztaló létezik.

Vallásos neveltetésemet az unitárius egyházban kaptam, ahol a világ összes vallását tanulmányoztuk. Hat hónapot töltöttünk, mondjuk, a buddhizmussal, és buddhista istentiszteletekre jártunk, elolvastuk a könyveiket, és beszélgetőcsoportokat tartottunk a vezetőikkel. Aztán átváltottunk egy másik vallásra, például a judaizmusra. A fő téma a "sok út vezet az igazsághoz" volt, a tolerancia és a transzcendencia mellett. Ez utóbbi gondolat azt jelentette, hogy a hagyományok közötti látszólagos ellentmondások feloldásához nem kell eldönteni, hogy az egyiknek igaza van, a másiknak pedig nincs. Az igazságot csak úgy lehet felfedezni, ha olyan magyarázatot találunk, amely felülírja-transzcendálja a látszólagos különbségeket, különösen az értelem és a cél alapvető kérdéseire.

Így oldom fel a tudatosság és a fizikai világ nyugati-keleti megosztottságát. Véleményem szerint mindkét nézőpontnak igaznak kell lennie.

Egyrészt ostobaság a fizikai világot tagadni. Még ha szimulációban élünk is, ahogy azt Nick Bostrom svéd filozófus feltételezi, a valóság mégis egy olyan fogalmi szint, amely számunkra valóságos. Ha elfogadjuk a fizikai világ létezését és a benne végbement evolúciót, akkor beláthatjuk, hogy a tudatos entitások ebből fejlődtek ki.

Másfelől a keleti szemlélet - miszerint a tudatosság alapvető és az egyetlen igazán fontos valóságot képviseli.

-szintén nehéz tagadni. Gondoljunk csak arra, hogy milyen nagyra becsüljük a tudatos személyeket a tudattalan dolgokkal szemben. Az utóbbiakat nem tekintjük belső értéknek, csak annyiban, amennyiben képesek befolyásolni a tudatos személyek szubjektív élményét. Még ha a tudatot egy komplex rendszer emergens tulajdonságának tekintjük is, nem foglalhatunk állást úgy, hogy az csak egy újabb tulajdonság (az "emésztés" és a "szoptatás" mellett, hogy John Searle-t idézzem). Azt képviseli, ami valóban fontos.

A "szellemi" szót gyakran használják a végső jelentőségű dolgok jelölésére. Sokan nem szívesen használják a spirituális vagy vallási hagyományokból származó ilyen

terminológiát, mert az olyan hiedelemkészleteket feltételez, amelyeket ők nem feltétlenül vallanak magukénak. Ha azonban levetkőzzük a vallási hagyományok misztikus bonyolultságát, és egyszerűen úgy tekintjük a "spirituális" kifejezést, mint ami az emberek számára valami mély értelmű dolgot jelent, akkor a tudatosság fogalma illik a képbe. A végső spirituális értéket tükrözi. Valójában magát a "szellemet" gyakran használják a tudatosság megjelölésére. Az evolúciót tehát spirituális folyamatnak tekinthetjük, mivel spirituális lényeket hoz létre, vagyis olyan entitásokat, amelyek tudatosak. Az evolúció a nagyobb komplexitás, a nagyobb tudás, a nagyobb intelligencia, a nagyobb szépség, a nagyobb kreativitás és a transzcendensebb érzelmek, például a szeretet kifejezésére való képesség felé is halad. Ezek mind olyan leírások, amelyeket az emberek Isten fogalmára használtak, bár Istent úgy írják le, hogy e tekintetben nincsenek korlátai.

Az emberek gyakran fenyegetve érzik magukat azoktól a vitáktól, amelyek azt a lehetőséget vetik fel, hogy egy gép tudatos lehet, mivel az ilyen irányú megfontolásokat a tudatos emberek szellemi értékének becsméréseként értékelik. Ez a reakció azonban a gép fogalmának félreértését tükrözi. Az ilyen kritikusok az általuk ma ismert gépek alapján közelítik meg a kérdést, és bármennyire is lenyűgözővé válnak, egyetértek azzal, hogy a technológia mai példái még nem méltóak arra, hogy tudatos lényekként tiszteljük őket. Az én előrejelzésem az, hogy ezek a gépek megkülönböztethetetlenek lesznek a biológiai emberektől, akiket tudatos lényeknek tekintünk, és ezért osztozni fognak abban a szellemi értékben, amelyet a tudatosságnak tulajdonítunk. Ez nem az emberek lebecsülése; inkább a jövő (egy-egy) gépei iránti megértésünk felértékelődése. Valószínűleg más terminológiát kellene alkalmaznunk ezekre az entitásokra, mivel másfajta gépek lesznek.

Valóban, ahogy most belenézünk az agyba és dekódoljuk a mechanizmusait, olyan módszereket és algoritmusokat fedezünk fel, amelyeket nemcsak megérteni, hanem újraalkotni is tudunk - "egy malom alkatrészei nyomják egymást", hogy Gottfried Wilhelm Leibniz (1646-1716) német matematikust és filozófust idézzem, amikor az agyról írt. Az emberek már most is szellemi gépezetek. Sőt, olyan szorosan össze fogunk olvadni az általunk létrehozott eszközökkel, hogy az ember és a gép közötti különbség addig fog elmosódni, amíg a különbség el nem tűnik. Ez a folyamat már javában zajlik, még akkor is, ha a legtöbb gép, amely meghosszabbít minket, még nincs a testünkben és az agyunkban.

Free Wi I

A tudatosság egyik központi aspektusa az előretétekintés képessége, amit "előrelátásnak" nevezünk. Ez a képesség a tervezésre, és társadalmi értelemben egy olyan forgatókönyv felvázolására szolgál, ami arról szól, hogy mi fog történni, vagy mi történhet olyan társadalmi interakciókban, amelyek még nem történtek meg.... Ez egy olyan rendszer, amelynek segítségével javítjuk az esélyeinket arra, hogy olyan dolgokat tegyünk, amelyek a saját érdekeinket képviselik.... Azt javaslom, hogy a "szabad akarat" a látszólagos képességünk arra, hogy ezek közül a leghasznosabbnak vagy legmegfelelőbbnek tűnő dolgokat válasszuk ki és cselekedjünk, és hogy

ragaszkodjunk ahhoz az elképzeléshez, hogy ezek a döntések a sajátjaink.

-Richard D. Alexander

Mondjuk azt, hogy a növény nem tudja, mit csinál, pusztán azért, mert nincs szeme, füle vagy agya? Ha azt mondjuk, hogy mechanikusan, és csakis mechanikusan cselekszik, nem kell-e elismernünk, hogy számos más, látszólag nagyon is tudatos cselekvés is mechanikus? Ha számunkra úgy tűnik, hogy a növény mechanikusan öl meg és eszik meg egy legyet, nem tűnhet-e úgy a növény számára, hogy az embernek mechanikusan kell megölnie és megennie egy birkát?

-Samuel Butler, 1871

Vajon az agy, amely feltűnően kettős szerkezetű, kettős szerv, "látszólag kettéválasztott, de mégis egy egység a kettéválasztásban"?

-Henry Maudsley⁸

A redundancia, mint már megtanultuk, a neokortex egyik legfontosabb stratégiája. Az agyban azonban a redundanciának van egy másik szintje is, mégpedig az, hogy a bal és a jobb agyfélteke, bár nem azonosak, nagyrészt azonosak. Ahogyan a neokortex bizonyos régiói általában bizonyos típusú információk feldolgozásával végzik, a féltekék is bizonyos mértékig specializálódnak - például a bal félteke jellemzően a verbális nyelvért felelős. De ezek a feladatmegosztások is átrendeződhetnek, olyannyira, hogy akár csak az egyik féltekével is képesek vagyunk életben maradni és valamennyire normálisan működni. Stella de Bode és Susan Curtiss amerikai neuropszichológus kutatók negyvenkilenc olyan gyermekről számoltak be, akik hemiszféréktómián (agyuk felének eltávolításán) estek át - ez egy olyan extrém műtét, amelyet olyan betegeken végeznek, akiknek életveszélyes, csak az egyik féltekéjüket érintő rohamos rendellenességük van. Az eljáráson átesettek közül néhányan hiányosságokkal maradnak, de ezek a hiányosságok specifikusak, és a betegek személyisége viszonylag normális. Sokan közülük boldogulnak, és a megfigyelők számára nem tűnik fel, hogy csak fél agyuk van. De Bode és Curtiss olyan bal félteke-eltávolított gyermekekről ír, akik "a "nyelvi" félteke eltávolítása ellenére figyelemre méltóan jó nyelvet fejlesztenek".² Leírnak egy ilyen diákot, aki elvégezte a főiskolát, diplomát szerzett, és átlagon felüli eredményt ért el az IQ-teszteken. A vizsgálatok minimális hosszú távú hatást mutattak ki az általános kognícióra, a memóriára, a személyiségre és a humorérzékre.¹⁰ Egy tanulmányban 2007-amerikai kutatók

Shearwood McClelland és Robert Maxwell hasonló hosszú távú pozitív eredményeket mutatott ki felnőtteknél.¹¹ Egy tízéves német kislányról, aki csak az agyának felével született, szintén beszámoltak, hogy teljesen normális. Sőt, az egyik szemére szinte tökéletesen lát, míg a féltekeeltávolításon átesett betegek a műtét után azonnal elveszítik látómezőjük egy részét.¹² Lars Muckli skót kutató így nyilatkozott: "Az agynak elképesztő képességekkel rendelkezik.

plaszticitást, de egészen meglepődünk, amikor láttuk, hogy ennek a lánynak az egyetlen agyféltekéje milyen jól alkalmazkodott a hiányzó félteke pótlásához."

Bár ezek a megfigyelések kétségtelenül alátámasztják a neokortex plasticitásának gondolatát, érdekesebb következményük az, hogy úgy tűnik, mindannyiunknak két agyunk van, nem egy, és mindkét agyvelővel elég jól boldogulunk. Ha elveszítjük az egyiket, akkor elveszítjük az ott tárolt egyedi agykérgi mintákat, de mindkét agy önmagában meglehetősen teljes. Tehát mindkét agyféltekének saját tudata van? Lehet érvelni amellett, hogy ez a helyzet.

Gondoljunk csak a hasadt agyú betegekre, akiknek még mindig megvan mindkét agyféltekéjük, de a köztük lévő csatornát elvágták. A corpus callosum egy körülbelül egymillió 250 axonból álló köteg, amely összeköti a bal és a jobb agyféltekét, és lehetővé teszi, hogy kommunikáljanak és koordináljanak egymással. Ahogyan két ember szorosan kommunikálhat egymással, és egyetlen döntéshozóként működhet, miközben különálló és egész egyéniségek maradnak, a két agyfélteke egységként működhet, miközben függetlenek maradnak.

Ahogy a kifejezés is mutatja, a hasadt agyú betegeknél a corpus callosum elvágódott vagy megsérült, így gyakorlatilag két működőképes agyuk van, amelyek között nincs közvetlen kommunikációs kapcsolat. Michael Gazzaniga (született 1939-ben) amerikai pszichológus kutató kiterjedt kísérleteket végzett azzal kapcsolatban, hogy a hasadt agyú betegeknél az egyes agyféltekék mit gondolnak.

A kettéosztott agyú betegeknél a bal félteke általában a jobb látómezőt látja, és fordítva. Gazzaniga és munkatársai egy hasadtagyú betegnek egy csirkekarmot ábrázoló képet mutattak a jobb látómezőbe (amelyet a bal félteke látott), és egy havas jelenetet a bal látómezőbe (amelyet a jobb félteke látott). Ezután egy képgyűjteményt mutatott úgy, hogy mindkét agyfélteke láthatta őket. Megkérte a páciens, hogy válasszon ki egyet a képek közül, amelyik jól illik az első képhez. A páciens bal keze (amelyet a jobb félteke irányított) egy lapát képére mutatott, míg a jobb keze egy csirke képére. Eddig minden rendben - a két félteke egymástól függetlenül és értelmesen működött. "Miért ezt választottad?" Gazzaniga megkérdezte a páciens, aki verbálisan válaszolt (a bal félteke beszédközpontja által vezérelve): "A csirkekárom nyilvánvalóan a csirkéhez illik". De aztán a páciens lenézett, és amikor észrevette, hogy a bal keze az ásóra mutat, azonnal megmagyarázta ezt (ismét a bal félteke által vezérelt beszédközpontjával): "és a csirkeól kitakarításához kell egy ásó".

Ez egy konfabuláció. A jobb félteke (amely a bal kart és kezét irányítja) helyesen mutat a lapátra, de mivel a bal félteke (amely a verbális választ irányítja) nincs tudatában a hónap, konfabulál egy magyarázatot, de nincs tudatában annak, hogy konfabulál. Felelősséget vállal egy olyan cselekedetért, amelyről soha nem döntött, és amelyet soha nem is hajtott végre, de azt hiszi, hogy megtette.

Ez azt jelenti, hogy a két agyféltekés páciens mindkét agyféltekéje saját tudattal rendelkezik. Úgy tűnik, a féltekék nincsenek tudatában annak, hogy a testüket ténylegesen két agy irányítja, mert megtanulnak egymással összehangolódni, és döntéseik kellően összehangoltak és következetesek ahhoz, hogy mindegyikük azt gondolja, hogy a másik félteke döntései a sajátjai.

Gazzaniga kísérlete nem bizonyítja, hogy egy normális, működő corpus callosummal rendelkező egyénnek két tudatos félagya van, de utal erre a lehetőségre. Bár a corpus

callosum lehetővé teszi a két agyfélteke hatékony együttműködését, ez nem feltétlenül jelenti azt, hogy nem különálló elmékről van szó. Mindkét agyfélteke elhitetheti, hogy minden döntést ő hozott, mert mindegyik elég közel állna ahhoz, amit mindegyik egyedül is eldöntött volna, és végül is nagy befolyása van minden egyes döntésre (azáltal, hogy a corpus callosumon keresztül együttműködik a másik féltekével). Tehát mind a két elme számára úgy tűnne, mintha ő irányítana.

Hogyan tesztelnéd azt a feltételezést, hogy mindketten tudatosak? Megvizsgálhatnánk őket a tudat neurológiai korrelátumai szempontjából, és Gazzaniga pontosan ezt tette. Kísérletei azt mutatják, hogy mindkét félteke önálló agyként működik. A konfabuláció nem korlátozódik az agyféltekékre; mindannyian rendszeresen csináljuk. Minden félteke körülbelül olyan intelligens, mint egy ember, tehát ha elhisszük, hogy az emberi agy tudatos, akkor arra kell következtetnünk, hogy minden félteke önállóan tudatos. Felmérhetjük a neurológiai korrelátumokat, és elvégezhetjük saját gondolatkísérleteinket (például figyelembe véve, hogy ha két agyfélteke működő corpus callosum nélkül két különálló tudatot alkot, akkor ugyanez igaz kell, hogy legyen két olyan agyféltekére is, amelyek között működő kapcsolat van), de minden kísérlet a tudat közvetlenebb kimutatására az egyes agyféltekékben ismét szembesít minket a tudatosság tudományos tesztjének hiányával. De ha megengedjük, hogy az agy mindkét féltekéje tudatos, akkor elismerjük-e, hogy a neokortexben zajló úgynevezett tudattalan tevékenység (amely a tevékenység túlnyomórésztét teszi ki) is rendelkezik önálló tudattal? Vagy talán több is van?

Marvin Minsky az agyat az "elme társadalmának" nevezi.¹³ Egy másik osztott agyú kísérletben a kutatók a "csengő" szót mutatták a jobb agyféltekének, a "zene" szót pedig a bal agyféltekének. A páciens megkérdezték, hogy melyik szót látja. A bal félteke által irányított beszédközpont azt mondta, hogy "zene". Ezután a kísérleti személynek egy képcsoportot mutattak, és megkérték, hogy mutasson rá arra a képre, amelyik a legközelebb áll az imént mutatott szóhoz. A jobb félteke által kontrollált karja a harangra mutatott. Amikor megkérdezték, miért mutatott a harangra, a bal félteke által vezérelt beszédközpontja azt válaszolta: "Hát, zene, utoljára akkor hallottam zenét, amikor a harangok kint dörömböltek". Ezt a magyarázatot annak ellenére adta meg, hogy más képek közül is választhatott, amelyek sokkal közelebb álltak a zenéhez.

Ez megint csak konfabuláció. A bal agyfélteke úgy magyaráz meg egy olyan döntést, mintha a sajátja lenne, amelyet soha nem hozott és nem hajtott végre. Nem azért teszi ezt, hogy egy barátját (vagyis a másik féltekéjét) fedezze - valóban azt hiszi, hogy a döntés a sajátja volt.

Ezek a reakciók és döntések érzelmi reakciókra is kiterjedhetnek. Megkérdezték egy tizenéves, osztott agyú páciens - így mindkét agyfélteke hallotta -, hogy "Ki a kedvenced...", majd a "barátnő" szót csak a jobb agyfélteke számára táplálták a bal fülön keresztül. Gazzaniga arról számol be, hogy az alany elpirult és zavartan viselkedett, ami egy tinédzsernek megfelelő reakció, amikor a barátnőjéről kérdezik. A bal félteke által vezérelt beszédközpont azonban azt jelentette, hogy nem hallotta a szót, és pontosítást kért: "A kedvenc micsodám?" Amikor ismét megkérték, hogy válaszoljon a kérdésre, ezúttal írásban, a jobb félteke által vezérelt bal kéz leírta a barátnője nevét.

Gazzaniga tesztjei nem gondolatkísérletek, hanem tényleges elmekísérletek. Miközben érdekes perspektívát kínálnak a tudatosság kérdéséhez, még közvetlenebbül szólnak a szabad akarat kérdéséhez. Mindegyik esetben az egyik félteke azt hiszi, hogy olyan döntést hozott, amelyet valójában soha nem hozott meg. Mennyire igaz ez a mindennapi döntéseinkre?

Tekintsük egy tízéves epilepsziás nöbeteg esetét. Itzhak Fried idegsebész éber állapotban végzett agyműtétet (ami azért lehetséges, mert az agyban nincsenek fájdalomreceptorok).¹⁴ Valahányszor a neokortex egy bizonyos pontját stimulálta, a lány nevetni kezdett. A sebészcsapat először azt hitte, hogy valamiféle nevetőreflexet váltanak ki, de hamar rájöttek, hogy a humor tényleges érzékelését váltják ki. Nyilvánvalóan megtalálták a neokortex egy pontját - nyilvánvalóan több is van belőle -, amely felismeri a humor érzékelését.

Nem csak nevetett - valóban viccesnek találta a helyzetet, még akkor is, ha valójában semmi sem változott a helyzetben azon kívül, hogy a neokortexének ezt a pontját stimulálták. Amikor megkérdezték tőle, hogy miért nevet, nem a "Ó, nincs különösebb oka" vagy a "Csak stimulálták az agyamat" típusú válaszokat adta, hanem rögtön egy okot mondott. Rámutatott valamire a szobában, és megpróbálta megmagyarázni, hogy az miért volt vicces. "Olyan viccesek vagytok, ahogy ott álltok" - volt egy tipikus megjegyzés.

Úgy tűnik, nagyon szívesen magyarázzuk és racionalizáljuk tetteinket, még akkor is, ha valójában nem mi hoztuk meg az őket kiváltó döntéseket. Mennyire vagyunk tehát felelősek a döntéseinkért? Nézzük meg Benjamin Libet (1916- 2007) fiziológiaprofesszor kísérleteit a Davis-i Kaliforniai Egyetemen. Libet a résztvevőket egy időzítő előtt ültette, fejbőrükre EEG-elektrodákat erősített. Egyszerű feladatok elvégzésére utasította őket, például egy gomb megnyomására vagy egy ujj mozgatására. A résztvevőket arra kérte, hogy jegyezzék fel az időmérőn azt az időpontot, amikor "először tudatosul bennük a cselekvés kívánsága vagy késztetése". A tesztek azt mutatták, hogy az alanyok mindössze 50 milliszekundumos hibahatárt mutattak ki ezeknél az értékeknél. Emellett átlagosan körülbelül 2 milliszekundumot00 mértek aközött az idő között, amikor az alanyok a cselekvés iránti késztetés tudatosítását jelentették, és a tényleges cselekvés között.¹⁵

A kutatók az alanyok agyából érkező EEG-jeleket is megvizsgálták. A motoros kéreg (amely a cselekvés végrehajtásáért felelős) cselekvés elindításában részt vevő agyi aktivitás valójában átlagosan körülbelül 500 milliszekundummal a feladat végrehajtása előtt történt. Ez azt jelenti, hogy a motoros kéreg körülbelül egy harmadmásodperccel azelőtt készült fel a feladat végrehajtására, hogy az alany egyáltalán tudatában lett volna annak, hogy meghozta az erre vonatkozó döntést.

A Libet-kísérletek következményeit hevesen vitatták. Maga Libet arra a következtetésre jutott, hogy a döntéshozatali tudatosságunk illúzióknak tűnik, hogy "a tudatosság kívül esik a

hurok." Daniel Dennett filozófus így kommentálta: "A cselekvés eredetileg az agy valamelyik részében indul el, és már repülnek is a jelek az izmok felé, útközben megállva, hogy elmondja neked, a tudatos szereplőnek, mi történik (de mint minden jó hivatalnok, hagyja, hogy te, a bambuló elnök, fenntartsd azt az illúziót, hogy te kezdted az egészet)." Dennett ugyanakkor megkérdőjelezte a kísérlet által rögzített időzítéseket, alapvetően azzal érvelve, hogy az alanyok nem biztos, hogy valóban tudatában vannak annak, hogy mikor válik tudatossá a cselekvésre vonatkozó döntés.¹⁶ Elgondolkozhatunk ezen: Ha az alany nincs tudatában annak, hogy mikor tudatosul benne a döntés meghozatala, akkor ki az? De a pont tulajdonképpen jól meg van fogalmazva - ahogyan azt korábban tárgyaltam -, hogy mi az, aminek tudatában vagyunk, korántsem egyértelmű.

Vilayanur Subramanian "Rama" Ramachandran (született 1951-ben) indiai-amerikai idegkutató kicsit másképp magyarázza a helyzetet. Tekintettel arra, hogy a neokortexben mintegy 30 milliárd neuronunk van, mindig sok minden történik ott, és tudatosan csak nagyon keveset veszünk észre belőle. A döntések, kicsik és nagyok, folyamatosan feldolgozásra kerülnek a neokortexben, és a javasolt megoldások felbukkannak a tudatos tudatunkban. Ramachandran azt javasolja, hogy a szabad akarat helyett inkább beszéljünk "szabad akaratról" - vagyis arról a képességről, hogy a neokortexünk nem tudatos részei által javasolt megoldásokat elutasíthatjuk.

Vegyük a katonai hadjárat analógiáját. A hadsereg tisztviselői ajánlást készítenek az elnöknek. Mielőtt megkapnák az elnök jóváhagyását, olyan előkészítő munkát végeznek, amely lehetővé teszi a döntés végrehajtását. Egy adott pillanatban a javasolt döntést bemutatják az elnöknek, aki jóváhagyja azt, és ezután megkezdődik a küldetés további része. Mivel az ezen analógia által képviselt "agy" magában foglalja a neokortex tudattalan folyamatait (vagyis az elnök alá tartozó tisztviselőket), valamint annak tudatos folyamatait (az elnököt), a hivatalos döntés meghozatalát megelőzően neurális tevékenységet és tényleges cselekvéseket is látnánk. Egy adott helyzetben mindig vitákba bocsátkozhatunk arról, hogy az elnök alatt álló tisztviselők valójában mekkora mozgásteret adtak az elnöknek egy ajánlás elfogadására vagy elutasítására, és bizonyára az amerikai elnökök mindkettőt megtették. De nem kellene meglepődnünk azon, hogy a mentális aktivitás, még a motoros kéregben is, már azelőtt elkezdődik, hogy tudatában lennénk annak, hogy döntést kell hozni.

Amit a Libet-kísérletek kiemelnek, az az, hogy az agyunkban sok olyan tevékenység van a döntéseink hátterében, amely nem tudatos. Azt már eddig is tudtuk, hogy a neokortexben zajló folyamatok nagy része nem tudatos; ezért nem meglepő, hogy cselekedeteink és döntéseink egyaránt származnak tudattalan és tudatos tevékenységből. Fontos ez a megkülönböztetés? Ha döntéseink mindkettőből erednek, számít-e, hogy a tudatos részeket kiválogatjuk-e a tudattalanból? Nem arról van-e szó, hogy mindkét aspektus képviseli az agyunkat? Nem mi vagyunk végső soron felelősek mindazért, ami az agyunkban zajlik? "Igen, lelöttem az áldozatot, de nem vagyok felelős, mert nem figyeltem" valószínűleg gyenge védekezés. Még ha vannak is olyan szűk jogi indokok, amelyek alapján egy személy nem vonható felelősségre a döntéseiért, általában minden döntésünkért felelősségre vonnak bennünket.

A fent idézett megfigyelések és kísérletek gondolat-kísérletek a szabad akarat

kérdésében, amely témát, akárcsak a tudat témáját, Platon óta vitatják. Maga a "szabad akarat" kifejezés a tizenharmadik századból származik, de mit is jelent pontosan?

A Merriam-Webster szótár meghatározása szerint "az emberek szabadsága, hogy olyan döntéseket hozzanak, amelyeket nem előzetes okok vagy isteni beavatkozás határoz meg". Észrevehetitek, hogy ez a definíció reménytelenül körkörös: "A szabad akarat szabadság...." Ha eltekintünk attól az elképzeléstől, hogy az isteni beavatkozás szemben áll a szabad akarral, van egy hasznos elem ebben a definícióban, mégpedig az, hogy a döntést "nem [határozzák meg] előzetes okok". Erre mindjárt visszatérek.

A *Stanford Encyclopedia of Philosophy* szerint a szabad akarat "a racionális ágensek azon képessége, hogy különböző alternatívák közül választanak egy cselekvési irányt". E definíció szerint egy egyszerű számítógép is képes a szabad akaratra, tehát kevésbé hasznos, mint a szótári definíció.

A Wikipédia valójában egy kicsit jobb. A szabad akaratot úgy definiálja, mint "az ágensek azon képességét, hogy bizonyos fajta kényszerektől mentes döntéseket hozzanak..... Az uralkodó kényszer a... determinizmus volt." Ismét a "szabad" körszót használja a szabad akarat meghatározásakor, de megfogalmazza azt, amit a szabad akarat legfőbb ellenségének tartanak: a *determinizmust*. Ebben a tekintetben a fenti Merriam-Webster-féle meghatározás tulajdonképpen hasonló, mivel olyan döntésekre utal, amelyeket "nem határoznak meg előzetes okok".

Mit értünk tehát determinizmus alatt? Ha beírom a "2 + 2"-t egy számológépbe, és az "4"-et mutat, mondhatom-e, hogy a számológép szabad akaratot mutatott, amikor úgy döntött, hogy a "4"-et mutatja? Senki sem fogadná el ezt a szabad akarat demonstrációjaként, mert a "döntést" a számológép belső mechanizmusai és a bevétel előre meghatározták. Ha egy bonyolultabb számítást teszek be, akkor is ugyanarra a következtetésre jutunk a szabad akarat hiányát illetően.

Mi a helyzet Watsonnal, amikor válaszol egy *Jeopardy!* kérdésre? Bár a számológépnél jóval összetettebbek a döntései, nagyon kevés megfigyelő, ha egyáltalán van olyan, aki szabad akaratot tulajdonítana a döntéseinek. Egyetlen ember sem tudja pontosan, hogyan működik az összes programja, de azonosítani tudunk egy olyan embercsoportot, akik együttesen képesek leírni az összes módszerét. Ennél is fontosabb, hogy a kimenetét (1) az összes programja határozza meg a lekérdezés pillanatában, (2) maga a lekérdezés, (3) a belső paramétereinek állapota, amelyek befolyásolják a döntéseit, és (4) a több billió bájtnyi tudásbázis, beleértve az enciklopédiákat is. E négy információkategória alapján határozza meg a kimenetét. Feltételezhetnénk, hogy ha ugyanazt a lekérdezést mutatjuk be, mindig ugyanazt a választ kapjuk, de Watson úgy van programozva, hogy tanuljon a tapasztalataiból, így fennáll a lehetősége annak, hogy a későbbi válaszok eltérőek legyenek. Ez azonban nem mond ellent ennek az elemzésnek; inkább csak a 3. pont, a döntéseit irányító paraméterek változását jelenti.

Tehát miben különbözik pontosan az ember Watson-tól, hogy az embernek szabad akaratot tulajdonítunk, de a számítógépes programnak nem? Több tényezőt is azonosíthatunk. Bár Watson jobb *Jeopardy!* játékos, mint a legtöbb, ha nem az összes ember, ennek ellenére közel sem olyan összetett, mint az emberi neokortex. Watson ugyan rengeteg tudással rendelkezik, és hierarchikus módszereket használ, de hierarchikus

gondolkodásának komplexitása még mindig jóval kisebb, mint egy emberé. A különbség tehát egyszerűen a hierarchikus gondolkodás komplexitásának mértékében van? Lehet érvelni amellett, hogy a kérdés valóban erre vezethető vissza. A tudatosság kérdésének tárgyalásakor megjegyeztem, hogy a saját hitem szerint egy olyan számítógépet, amely átmegy egy érvényes Turing-teszten, tudatosnak tekintenék. A legjobb chatbotok ma még nem képesek erre (bár folyamatosan fejlődnek), így a tudatossággal kapcsolatos következtetésem az entitás teljesítményszintjének kérdése. Talán ugyanez igaz arra is, hogy szabad akaratot tulajdonítok neki.

A tudatosság valóban az egyik filozófiai különbség az emberi agy és a mai szoftverprogramok között. Az emberi agyat tudatosnak tekintjük, míg a szoftverprogramoknak ezt még nem tulajdonítjuk. Ez az a tényező, amit keresünk, ami a szabad akarat alapja?

Egy egyszerű elmekísérlet azt bizonyítja, hogy a tudatosság valóban a szabad akarat létfontosságú része. Gondoljunk egy olyan helyzetre, amelyben valaki úgy hajt végre egy cselekvést, hogy nincs tudatában annak, hogy ezt teszi - a cselekvést teljes egészében az adott személy agyának nem tudatos tevékenysége hajtja végre. Ezt a szabad akarat megnyilvánulásának tekintenék? A legtöbb ember nemmel válaszolna. Ha a cselekedet káros lenne, valószínűleg akkor is felelősnek tartanánk az illetőt, de keresnénk néhány olyan közelmúltbeli tudatos cselekedetet, ami miatt az illető tudatos tudatosság nélkül hajtott végre cselekvéseket, például ha egy itallal többet ivott volna, mint amennyit kellett, vagy ha egyszerűen csak nem képezte magát megfelelően arra, hogy tudatosan mérlegelje döntéseit, mielőtt cselekedett volna.

Egyes kommentátorok szerint a Libet-kísérletek a szabad akarat ellen érveltek azzal, hogy rávilágítottak arra, hogy döntéseink nagy része nem tudatos. Mivel a filozófusok között ésszerű konszenzus van abban, hogy a szabad akarat tudatos döntéshozatalt feltételez, úgy tűnik, hogy ez a szabad akarat egyik előfeltétele. Sok megfigyelő számára azonban a tudatosság szükséges, de nem elégséges feltétel. Ha a döntéseink - tudatosak vagy nem tudatosak - előre meghatározottak, mielőtt meghozzuk őket, hogyan mondhatjuk, hogy a döntéseink szabadok? Ezt az álláspontot, amely szerint a szabad akarat és a determinizmus nem összeegyeztethető, inkompatibilizmusnak nevezik. Carl Ginet amerikai filozófus (szül. 1932) például úgy érvel, hogy ha a múltban, a jelenben és a jövőben bekövetkező események determináltak, akkor úgy tekinthető, hogy nincs befolyásunk felettük vagy következményeik felett. Látszólagos döntéseink és cselekedeteink egyszerűen ennek az előre meghatározott sorrendnek a részei. Ginet szerint ez kizárja a szabad akaratot.

A determinizmust azonban nem mindenki tartja összeegyeztethetetlennek a szabad akarat fogalmával. A kompatibilisták lényegében azzal érvelnek, hogy az ember szabadon dönthet arról, hogy mit akar, még akkor is, ha az, hogy mit dönt, determinált vagy determinált lehet. Daniel Dennett például azzal érvel, hogy bár a jelen állapotából meghatározható a jövő, a valóság az, hogy a világ olyan bonyolultan összetett, hogy nem tudhatjuk, mit hoz a jövő. Azonosíthatjuk azt, amit ő "elvárásoknak" nevez, és valóban szabadon végrehajthatunk olyan cselekedeteket, amelyek eltérnek ezektől az elvárásoktól. Azt kell mérlegelnünk, hogy döntéseink és cselekedeteink hogyan viszonyulnak ezekhez az elvárásokhoz, nem pedig egy elméletileg meghatározott jövőhöz, amelyet valójában nem

ismerhetünk. Dennett szerint ez elegendő a szabad akarathoz.

Gazzaniga szintén egy kompatibilista álláspontot fogalmaz meg: "Személyesen felelős ágensek vagyunk, és felelősségre kell vonnunk tetteinkért, még akkor is, ha egy determinált világban élünk." ¹⁷ Egy cinikus így értelmezheti ezt a nézetet: Nincs kontrollod a tetteid felett, de mi mégis téged hibáztatunk.

Egyes gondolkodók a szabad akarat gondolatát illúzióknak tekintik. David Hume skót filozófus (1711-1776) egyszerűen "verbális" dolognak nevezte, amelyet "hamis érzés vagy látszólagos tapasztalat" jellemez. ¹⁸ Arthur Schopenhauer német filozófus (1788-1860) azt írta, hogy "mindenki *a priori* azt hiszi magáról, hogy tökéletesen szabad, még az egyéni cselekedeteiben is, és azt hiszi, hogy minden pillanatban másfajta életet kezdhet..... De *utólag*, a tapasztalat révén, megdöbbenésére tapasztalja, hogy nem szabad, hanem alá van vetve a kényszernek, hogy minden elhatározása és elmélkedése ellenére sem változtat magatartásán, és hogy élete kezdetétől annak végéig éppen azt a jellemet kell megvalósítania, amelyet ő maga elítél." ¹⁹

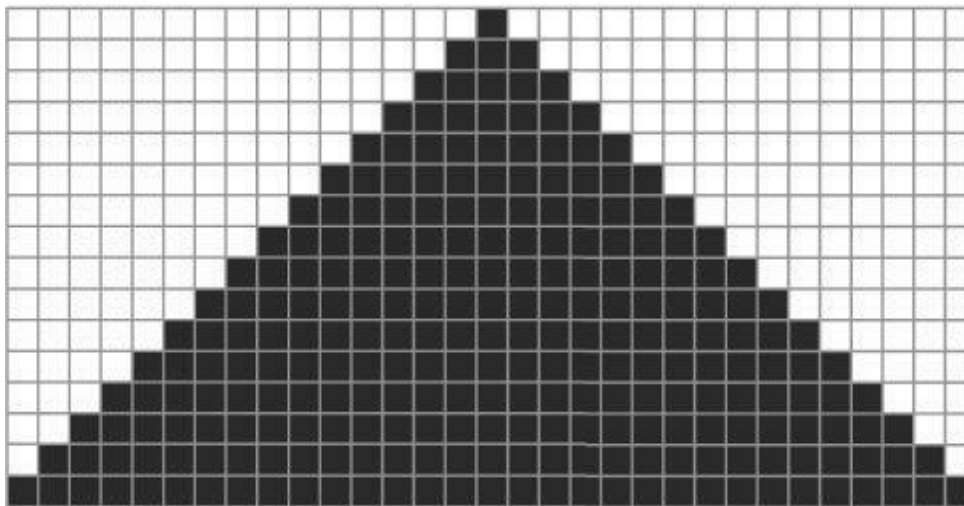
Itt több pontot is hozzátennék. A szabad akarat - és a felelősség fogalma, amely szorosan kapcsolódik hozzá - hasznos, sőt, létfontosságú a társadalmi rend fenntartásához, függetlenül attól, hogy a szabad akarat valóban létezik-e vagy sem. Ahogy a tudat egyértelműen létezik mémként, úgy a szabad akarat is. A létezésének bizonyítására vagy akár a meghatározására tett kísérletek reménytelenül körkörösé válhatnak, de a valóság az, hogy szinte mindenki hisz ebben az elképzelésben. Magasabb szintű neokortexünk igen jelentős részét szenteljük annak a koncepciónak, hogy szabad döntéseket hozunk, és felelősek vagyunk a tetteinkért. Függetlenül attól, hogy szigorú filozófiai értelemben ez igaz-e vagy egyáltalán lehetséges-e, a társadalom sokkal rosszabbul járna, ha nem hinnénk ilyen dolgokban.

Továbbá a világ nem feltétlenül determinált. Fentebb a kvantummechanika két nézőpontját tárgyaltam, amelyek a kvantummezőknek a megfigyelőhöz való viszonyát illetően különböznek egymástól. A megfigyelő-alapú perspektíva egyik népszerű értelmezése szerepet szán a tudatnak: A részecskék nem oldják fel kvantum kétértelműségüket, amíg egy tudatos megfigyelő meg nem figyeli őket. A kvantum események filozófiájában van egy másik megosztottság is, amely hatással van a szabad akaratról folytatott vitánkra, és amely a kérdés körül forog: A kvantum események determináltak vagy véletlenszerűek?

A kvantum esemény legáltalánosabb értelmezése az, hogy amikor a részecskét alkotó hullámfüggvény "összeomlik", a részecske helye konkrétvá válik. Nagyon sok ilyen eseményt tekintve lesz egy megjósolható eloszlás (ezért tekintik a hullámfüggvény valószínűségi eloszlásnak), de minden egyes ilyen, a hullámfüggvény összeomlásán áteső részecske felbontása véletlenszerű. Az ezzel ellentétes értelmezés determinisztikus: pontosabban, hogy van egy rejtett változó, amelyet nem tudunk külön-külön kimutatni, de amelynek értéke meghatározza a részecske helyzetét. A rejtett változó értéke vagy fázisa a hullámfüggvény összeomlásának pillanatában határozza meg a részecske helyzetét. A legtöbb kvantumfizikus látszólag a valószínűségi mező szerinti véletlenszerű felbontás gondolatát részesíti előnyben, de a kvantummechanika egyenletei lehetővé teszik egy ilyen

fekete cellával kezdünk középen, és a cellák fejlődését több generáción keresztül mutatjuk be (ahol minden sor, ahogy lefelé haladunk, egy új értékgenerációt jelent), a 222-es szabály eredményei így néznek ki:

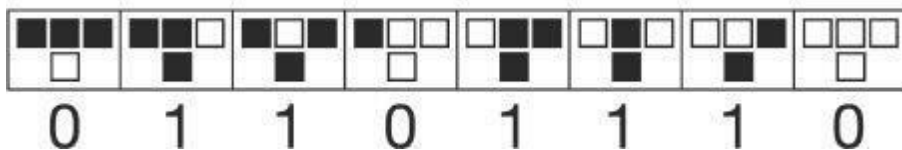
rule 222



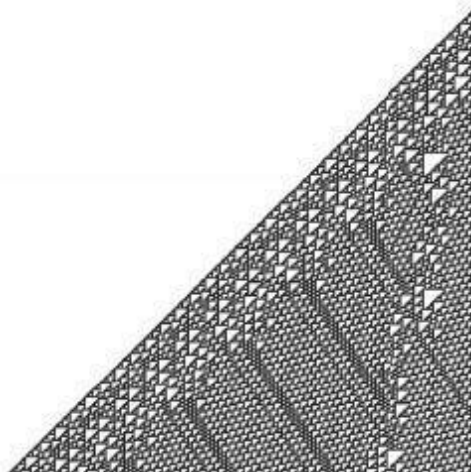
Egy automata egy szabályon alapul, és egy szabály határozza meg, hogy a cella fekete vagy fehér lesz-e az alapján, hogy a nyolc lehetséges minta közül melyik létezik az aktuális generációban. Így $2^8 = 256$ lehetséges szabály van. Dr. Wolfram felsorolta mind a 256 lehetséges ilyen automatát, és mindegyikhez Wolfram-kódot rendelt a következőktől az 0Érdekes255. módon ez a 256 elméleti automata nagyon különböző tulajdonságokkal rendelkezik. A Dr. Wolfram által I. osztálynak nevezett automaták, mint például a 222-es szabály, nagyon kiszámítható mintákat hoznak létre. Ha megkérdezném, hogy a szabály trillió trillió iterációja után mi a középső cella értéke, 222, könnyen válaszolhatnánk: fekete.

Sokkal érdekesebbek azonban a IV. osztályú automaták, amelyeket a 110. szabály szemléltet.

rule 110



Ennek az automatának több generációja így néz ki:



A szabályautomaták110 és általában a IV. osztályú automaták érdekessége, hogy az eredmények teljesen kiszámíthatatlanok. Az eredmények átmennek a legszigorúbb matematikai véletlenszerűségi teszteken, mégsem egyszerűen zajt generálnak: Vannak ismétlődő minták, de ezek furcsa és kiszámíthatatlan módon ismétlődnek. Ha megkérdezném, hogy mi egy adott cella értéke egy trillió trillió iteráció után, nem lenne mód arra, hogy megválaszoljuk ezt a kérdést anélkül, hogy ténylegesen lefuttatnánk ezt a gépet ennyi generáción keresztül. A megoldás egyértelműen meghatározott, mert ez egy nagyon egyszerű determinisztikus gép, de teljesen kiszámíthatatlan anélkül, hogy ténylegesen lefuttatnánk a gépet.

Dr. Wolfram elsődleges tézise az, hogy a világ egy nagy IV. osztályú celluláris automata. Azért kapta könyvének az *Egy újfajta tudomány* címet, mert ez az elmélet ellentétben áll a legtöbb más tudományos törvényszerűséggel. Ha van egy műhold, amely a Föld körül kering, megjósolhatjuk, hogy hol lesz öt év múlva anélkül, hogy a gravitáció vonatkozó törvényeinek felhasználásával végig kellene futnunk egy szimulált folyamat

minden egyes pillanatát, és meg tudnánk oldani, hogy hol lesz a műhold a messzi jövőbeni időpontokban. A IV. osztályú celluláris automaták jövőbeli állapotát azonban nem lehet megjósolni anélkül, hogy minden egyes lépést szimulálnánk az út során. Ha az univerzum egy óriási celluláris automata, ahogy Dr. Wolfram posztulálja, akkor nem létezne elég nagy számítógép - mivel minden számítógép az univerzum egy részhalmaza lenne.

-amely képes egy ilyen szimulációt lefuttatni. Ezért a világegyetem jövőbeli állapota teljesen megismerhetetlen, még akkor is, ha determinisztikus.

Így bár döntéseink determináltak (mivel testünk és agyunk egy determinisztikus univerzum része), mégis eredendően kiszámíthatatlanok, mivel egy IV. osztályú automatában élünk (és annak részei vagyunk). Egy IV. osztályú automata jövőjét nem tudjuk megjósolni, csak ha hagyjuk, hogy a jövő kibontakozzon. Dr. Wolfram számára ez elegendő ahhoz, hogy szabad akaratot tegyünk lehetővé.

Nem kell a világegyetemre tekintenünk, hogy meglássuk a jövő eseményeit, amelyek meghatározottak, mégis kiszámíthatatlanok. A Watsonon dolgozó tudósok egyike sem tudja megjósolni, hogy mit fog tenni, mert a program egyszerűen túlságosan összetett és változatos, és teljesítménye olyan tudáson alapul, amely túlságosan kiterjedt ahhoz, hogy bármely ember elsajátíthassa. Ha hiszünk abban, hogy az embereknek szabad akaratuk van, akkor ebből az következik, hogy a Watson vagy a Watson-szerű gépek jövőbeli változatainak is szabad akaratuk van.

A saját hitem az, hogy hiszek abban, hogy az embereknek szabad akaratuk van, és bár úgy teszek, mintha ez lenne a helyzet, nehezen találok olyan példákat a saját döntéseim között, amelyek ezt illusztrálják. Gondoljunk csak arra a döntésre, hogy megírom ezt a könyvet - ezt a döntést soha nem hoztam meg. Inkább a könyv ötlete döntötte el ezt helyettem. Általánosságban úgy érzem, hogy olyan ötletek foglya vagyok, amelyek úgy tűnik, hogy beültetik magukat a neokortexembe, és átveszik az irányítást. Mi a helyzet a házasságról szóló döntéssel, amelyet (egy másik emberrel együtt) harminchat évvel ezelőtt hoztam meg? Akkoriban a szokásos programot követtem: vonzódtam - és üldöztem - egy csinos lányt. Aztán szerelmes lettem. Hol van ebben a szabad akarat?

De mi van azokkal az apró döntésekkel, amelyeket minden nap hozok - például, hogy milyen szavakat választok a könyvemben? Egy üres virtuális papírlappal kezdek. Senki sem mondja meg, hogy mit tegyek. Nincs szerkesztő, aki a vállam fölé nézne. A döntéseim *teljesen* rajtam múlnak. Szabadon - teljesen *szabadon* - azt írok, amit akarok...

Uh, *grok*...

Grok? Oké, megcsináltam - végre alkalmaztam a szabad akaratomat. Azt akartam írni, hogy "akarom", de szabad elhatározásból valami teljesen váratlan dolgot írtam helyette. Talán ez az első alkalom, hogy sikerült tiszta szabad akaratot gyakorolnom.

Vagy nem.

Nyilvánvalónak kell lennie, hogy ez nem az akarat megnyilvánulása volt, hanem inkább egy pont illusztrálására tett kísérlet (és talán egy gyenge humorérzék).

Bár osztom Descartes meggyőződését, hogy tudatos vagyok, a szabad akarattal kapcsolatban nem vagyok olyan biztos. Nehéz kibújni Schopenhauer következtetése alól, miszerint "azt tehetsz, amit akarsz, de életed bármelyik pillanatában csak egy határozott

dolgot *akarhatsz*, és azon az egy dolgon kívül abszolút semmi más".²⁰ Ennek ellenére továbbra is úgy fogok cselekedni, mintha lenne szabad akaratom, és hiszek benne, amíg nem kell megmagyaráznom, hogy miért.

Identitás

Egy filozófus egyszer a következő álmot látta.

Először Arisztotelész jelent meg, és a filozófus azt kérdezte tőle: "Adnál nekem egy tizenöt perces kapszulavázlatot az egész filozófiádról?". A filozófus meglepetésére Arisztotelész egy kiváló fejtegetést adott neki, amelyben hatalmas mennyiségű anyagot sűrített bele mindössze tizenöt percbe. Ekkor azonban a filozófus felvetett egy bizonyos ellenvetést, amelyre Arisztotelész nem tudott válaszolni. Megzavarodva Arisztotelész eltűnt.

Aztán megjelent Platón. Ugyanez ismét megtörtént, és a filozófus Platónnal szembeni ellenvetése ugyanaz volt, mint az Arisztotelésszel szembeni ellenvetése. Platón erre sem tudott válaszolni, és eltűnt.

Aztán a történelem összes híres filozófusa megjelent egymás után, és a mi filozófusunk mindegyiket ugyanazzal az ellenvetéssel cáfolta. Miután az utolsó filozófus is eltűnt, filozófusunk azt mondta magában: "Tudom, hogy alszom és álmodom mindezt. Mégis megtaláltam az összes filozófiai rendszer egyetemes cáfolatát! Holnap, mire felébredek, valószínűleg elfelejtem, és a világ tényleg lemarad valamiről!" A filozófus vaskos erőfeszítéssel kényszerítette magát, hogy felébredjen, az íróasztalához sietett, és leírta az egyetemes cáfolatot. Aztán megkönnyebbült sóhajjal ugrott vissza az ágyba.

Másnap reggel, amikor felébredt, odament az íróasztalhoz, hogy megnézzze, mit írt. Azt írta: "Ezt mondod *te*". -Raymond Smullyan, David Chalmers idézi.²¹

Amire még inkább kíváncsi vagyok, mint arra, hogy tudatos vagyok-e vagy nem vagyok-e szabad akarátú, az az, hogy miért vagyok tudatában ennek a bizonyos személynek a tapasztalatainak és döntéseinek, aki könyveket ír, szeret túrázni és kerékpározni, táplálékkiegészítőket szed, és így tovább. A kézenfekvő válasz az lenne, hogy "Mert ez vagy te".

Ez a csere valószínűleg nem tautologikusabb, mint a fenti válaszaim a tudattal és a szabad akaratral kapcsolatos kérdésekre. De valójában van egy jobb válaszom arra, hogy miért kapcsolódik a tudatom ehhez a bizonyos személyhez: Azért, mert annak teremtettem magam.

Egy gyakori aforizma szerint: "Az vagy, amit megeszel". Még igazabb, ha azt mondjuk: "Az vagy, amit gondolsz". Amint azt már megbeszéltük, a neokortexemben lévő összes hierarchikus struktúra, amely meghatározza a személyiséget, a képességeimet és a tudásomat, a saját gondolataim és tapasztalataim eredménye. Az emberek, akikkel úgy

döntök, hogy kapcsolatba kerülök, valamint az ötletek és projektek, amelyekben részt veszek, mind elsődlegesen meghatározzák, hogy kivé válok. Ami azt illeti, az, hogy mit eszem, szintén a neokortexem által hozott döntéseket tükrözi. Egyelőre elfogadva a szabad akarat kettősségének pozitív oldalát, a saját döntéseim eredményezik azt, aki vagyok.

Függetlenül attól, hogy hogyan lettünk azok, akik vagyunk, mindannyiunkban él a vágy, hogy identitásunk megmaradjon. Ha nem lenne meg benned az akarat a fennmaradásra, akkor nem lennél itt, és nem olvasnád ezt a könyvet. Minden élőlénynek megvan ez a célja - ez az evolúció legfőbb meghatározója. Az identitás kérdését talán még nehezebb meghatározni, mint a tudatosságot vagy a szabad akaratot, de vitathatatlanul fontosabb. Végül is tudnunk kell, hogy mik vagyunk, ha meg akarjuk őrizni a létezésünket.

Gondoljunk erre a gondolat kísérletre: Ön a jövőben van, ahol a technológiák fejlettebbek, mint a maiak. Miközben Ön alszik, egy csoport letapogatja az agyát, és minden lényeges részletet rögzít. Lehet, hogy ezt az agyad hajszálereiben utazó, véresejt méretű letapogató gépekkel teszik, vagy valamilyen más alkalmas, nem invazív technológiával, de egy adott időpontban minden információ a birtokukban van az agyadról. Emellett felvesznek és rögzítenek minden olyan testi részletet is, amely tükrözheti az Ön lelkiállapotát, például az endokrin rendszerét. Ezt az "elmefájlt" egy olyan nem biológiai testbe helyezik, amely úgy néz ki és úgy mozog, mint te, és rendelkezik a szükséges finomsággal és hajlékonysággal ahhoz, hogy a te neveddel azonosuljon. Reggel értesülsz erről az áthelyezésről, és végignézed (talán észrevétlenül), ahogyan az elméd klónja, akit úgy fogunk hívni, hogy Te Te2., úgy beszél az életéről, mintha Te lennél, és elmeséli, hogyan fedezte fel még aznap reggel, hogy egy sokkal tartósabb új verziót kapott.

2.0 test. "Hé, nagyon tetszik ez az új test!" - kiáltja.

Az első kérdés, amit meg kell fontolnunk: tudatos vagy-e a You 2? Nos, úgy tűnik, hogy igen. Átmegy az általam korábban megfogalmazott teszten, mivel rendelkezik az érző, tudatos személyiség finom jelzéseivel. Ha te tudatos vagy, akkor te is az vagy. 2.

Szóval, ha eltünnél, senki sem venné észre. Ti ketten körbejárnátok, és azt állítanátok, hogy ti vagytok. Minden barátod és szeretted megelégedne a helyzettel, és talán örülne is, hogy most már tartósabb testtel és mentális szubsztrátummal rendelkezel, mint korábban. Talán a filozofikusabban gondolkodó barátaid kifejeznék aggodalmukat, de a legtöbb esetben mindenki boldog lenne, beleértve téged is, vagy legalábbis azt a személyt, aki meggyőzően azt állítja magáról, hogy te vagy.

Szóval már nincs szükségünk a régi testedre és agyadra, igaz? Nem baj, ha megszabadulunk tőle?

Valószínűleg nem fogsz belemenni ebbe. Jelezttem, hogy a vizsgálat nem volt invazív, tehát még mindig itt van, és még mindig magánál van. Ráadásul az identitástudatod még mindig nálad van, nem pedig Te 2-nél, még akkor is, ha Te azt hiszed², hogy ő a Te folytatásod. Lehet², hogy nem is tudatosul benned, hogy létezel, vagy valaha is léteztél. Valójában a Te 2 létezésének sem lennél tudatában, ha nem szóltunk volna róla.

A következtetésünk? Te 2 tudatos vagy, de más személy, mint te - Te 2 más identitással rendelkezik. Rendkívül hasonlít, sokkal jobban, mint egy egyszerű genetikai klón, mert osztozik az összes neokortikális mintázatodban és kapcsolatodban. Vagy inkább

azt kellene mondanom, hogy megosztotta ezeket a mintákat abban a pillanatban, amikor létrejött. Azon a ponton kezdtétek el a saját utatokat járni, neokortikális értelemben. Ti még mindig itt vagytok. Nem ugyanazokat a tapasztalatokat éled át, mint Te 2. A lényeg: Te 2 nem te vagy.

Oké, eddig minden rendben. Most nézzünk meg egy másik gondolat kísérletet - azt hiszem, hogy reálisabb a jövőre nézve. Ön aláveti magát egy olyan eljárásnak, amelynek során az agyának egy nagyon kis részét egy nem biológiai egységgel cserélik ki. Meg van győződve arról, hogy ez biztonságos, és különböző előnyökről számolnak be.

Ez nem is olyan távoli, hiszen a neurológiai és érzékszervi károsodásban szenvedő emberek esetében rutinszerűen alkalmazzák, mint például a Parkinson-kórban szenvedő embereknél alkalmazott neurális implantátumot és a süketeknél alkalmazott cochleáris implantátumot. Ezekben az esetekben a számítógépes eszközt a testen belül, de az agyon kívül helyezik el, mégis az agyba (vagy a cochleáris implantátumok esetében a hallóidegbe) csatlakoztatják. Véleményem szerint az a tény, hogy a tényleges számítógép fizikailag a tényleges agyon kívül helyezkedik el, filozófiailag nem jelentős: Gyakorlatilag az agyat egészítjük ki, és egy számítógépes eszközzel helyettesítjük azokat a funkcióit, amelyek már nem működnek megfelelően. A 2030-as években, amikor az intelligens számítógépes eszközök vérsajt méretűek lesznek (és ne feledjük, hogy a fehérvérsejtek elég intelligensek ahhoz, hogy felismerjék és leküzdjék a kórokozókat), noninvazív módon fogjuk őket bevezetni, nem lesz szükség műtétre.

Visszatérve a jövőbeli forgatókönyvünkhöz, elvégezted az eljárást, és ahogy ígérted, remekül működik - bizonyos képességeid javultak. (Talán jobb a memóriád.) Szóval még mindig te vagy? A barátaid biztosan így gondolják. Ön is így gondolja. Nincs jó érv arra, hogy hirtelen más ember lettél. Nyilvánvaló, hogy azért vetted alá magad a beavatkozásnak, hogy valamiben változást érsz el, de még mindig ugyanaz vagy, aki vagy. Az identitásod nem változott. Nem egy másik ember tudata vette át hirtelen a testét.

Oké, tehát az eredményeken felbátorodva úgy dönt, hogy újabb beavatkozásra kerül sor, ezúttal az agy egy másik régióját érintve. Az eredmény ugyanaz: némi javulást tapasztalsz a képességeidben, de még mindig te vagy.

Nyilvánvalónak kell lennie, hogy mire akarok kilyukadni. Folyamatosan további eljárásokat választasz, a folyamatba vetett bizalmad csak növekszik, míg végül az agyad minden részét megváltoztatod. Minden egyes alkalommal az eljárást gondosan úgy végezték, hogy megőrizzék az összes neokortikális mintázatodat és kapcsolatodat, így nem veszítetél semmit a személyiségedből, képességeidből vagy emlékeidből. Soha nem volt egy Te és egy Te 2; csak Te voltál. Senki, beleértve téged is, soha nem veszi észre, hogy megszűntél létezni. Valóban - itt vagy.

Következtetésünk: Még mindig létezel. Itt nincs dilemma. Minden rendben van.

Kivéve ezt: Te a fokozatos cserefolyamat után teljesen egyenértékű vagy az előző gondolat kísérletben szereplő 2. Veled (amit én a letapogatási és megállapítási forgatókönyvnek fogok nevezni). A fokozatos helyettesítés forgatókönyve után Te rendelkezél az összes neokortikális mintával és kapcsolattal, amelyekkel eredetileg is rendelkezted, csak nem biológiai szubsztrátumban, ami igaz a Te 2-re is a letapogatás és

megalapozás forgatókönyvében. A fokozatos helyettesítés forgatókönyve után rendelkezel néhány további képességgel és nagyobb tartóssággal, mint a folyamat előtt, de ez ugyanígy igaz a Te 2-re is a szkennelés-alapozás folyamatában.²

De arra a következtetésre jutottunk, hogy a You 2 *nem* te vagy. És ha te a fokozatos cserefolyamat után teljesen egyenértékű vagy a Te 2-vel a beolvasás és a megalapozás folyamata után, akkor te a fokozatos cserefolyamat után szintén nem lehetsz te.

Ez azonban ellentmond korábbi következtetésünknek. A fokozatos cserefolyamat több lépésből áll. Úgy tűnt, hogy ezek a lépések mindegyike megőrzi az identitást, ahogyan ma arra a következtetésre jutunk, hogy egy Parkinson-kóros betegnek ugyanaz az identitása, miután beültettek egy idegi implantátumot.²²

Éppen ez a fajta filozófiai dilemma vezet egyeseket arra a következtetésre, hogy ezek a csereforgatókönyvek soha nem fognak bekövetkezni (még akkor sem, ha már most is zajlanak). De gondoljunk csak bele: Életünk során természetes módon fokozatos cserefolyamaton megyünk keresztül. A testünkben lévő sejtek többsége folyamatosan cserélődik. (Őn éppen most cserélt ki 100 milliót közülük az utolsó mondat elolvasása során). A vékonybél belső bélrendszerének sejtjei körülbelül egy hét alatt cserélődnek le, akárcsak a gyomor védőbélése. A fehérvérsejtek élettartama típustól függően néhány naptól néhány hónapig terjed. A vérlemezkék körülbelül kilenc napig tartanak.

Az idegsejtek fennmaradnak, de szerveik és alkotó molekuláik egy hónapon belül átalakulnak.²³ A neuron mikrotubulusának felezési ideje körülbelül tíz perc; a dendritekben lévő aktin filamentumok körülbelül negyven másodpercig tartanak; a fehérvérjék, amelyek az energiát biztosítják a

a szinapszisok óránként cserélődnek; a szinapszisokban lévő NMDA-receptorok viszonylag hosszú élettartamúak, öt nap.

Tehát néhány hónap alatt teljesen kicserélődsz, ami hasonló a fent leírt fokozatos csere forgatókönyvéhez. Ugyanaz a személy vagy, aki néhány hónappal ezelőtt voltál? Bizonyára vannak különbségek. Talán tanultál néhány dolgot. De feltételezed, hogy az identitásod megmarad, hogy nem pusztulsz el és nem teremted újra folyamatosan.

Gondoljunk egy folyóra, mint amilyen az irodám mellett folyik. Ahogy most kinézek arra, amit az emberek Charles folyónak neveznek, vajon ugyanaz a folyó, amit tegnap láttam? Először is gondolkodjunk el azon, hogy mi is az a folyó. A szótár meghatározása szerint "egy nagy természetes áramló vízfolyás". E definíció szerint a folyó, amit most látok, teljesen más, mint tegnap volt. Minden egyes vízmolekulája megváltozott, és ez a folyamat nagyon gyorsan végbemegy. Diogenész Laertiosz görög filozófus írta a Krisztus utáni harmadik században, hogy "nem lehet kétszer ugyanabba a folyóba lépni".

De általában nem így tekintünk a folyókra. Az emberek szeretnek rájuk nézni, mert a folyamatosság és a stabilitás szimbólumai. Az általános nézet szerint a Charles folyó, amelyet tegnap néztem, ugyanaz a folyó, amelyet ma is látok. Az életünk nagyjából ugyanolyan. Alapvetően nem azok vagyunk, amik a testünket és az agyunkat alkotják. Ezek a részecskék lényegében ugyanúgy áramlanak rajtunk keresztül, mint ahogy a vízmolekulák áramlanak egy folyón. Olyan minta vagyunk, amely lassan változik, de stabilitással és folytonossággal rendelkezik, még akkor is, ha a mintát alkotó anyag gyorsan változik.

A nem biológiai rendszerek fokozatos bevezetése a testünkbe és az agyunkba csak egy újabb példája lesz a minket alkotó részek folyamatos cserélődésének. Ez éppúgy nem fogja megváltoztatni identitásunk folyamatosságát, mint a biológiai sejtjeink természetes cseréje. Történelmi, szellemi, társadalmi és személyes emlékeinket már most is nagyrészt kiszerveztük az eszközeinkbe és a felhőbe. Az eszközök, amelyekkel kapcsolatba lépünk, hogy hozzáférjünk ezekhez az emlékekhez, talán még nincsenek a testünkben és az agyunkban, de ahogy egyre kisebbek lesznek (és a technológiát évtizedenként körülbelül százás ütemben zsugorítjuk 3D térfogatban), úgy fognak oda kerülni. Mindenesetre hasznos hely lesz, ahová elhelyezhetjük őket - így nem fogjuk elveszíteni őket. Ha az emberek mégis lemondanak arról, hogy mikroszkopikus eszközöket helyezzenek el a testükben, az sem lesz baj, hiszen más módon is hozzáférhetnek majd az átható felhőintelligenciához. De visszatérünk a korábban bemutatott dilemmához. Ön a fokozatos csere időszaka után egyenértékű Ön 2-vel a szkennelés és beazonosítás forgatókönyvében, de úgy döntöttünk, hogy Ön 2 ebben a forgatókönyvben nem rendelkezik ugyanazzal az identitással, mint Ön. Hol maradunk tehát?

A nem biológiai rendszerek olyan képességének megbecsülését hagyja bennünk, amellyel a biológiai rendszerek nem rendelkeznek: a másolás, a biztonsági mentés és az újrateremtés képessége. Ezt rendszeresen megteesszük az eszközeinkkel. Amikor új okostelefonot veszünk, átmásoljuk az összes fájlunkat, így az nagyjából ugyanazzal a személyiséggel, képességekkel és emlékekkel rendelkezik, mint a régi okostelefon. Talán néhány új képességgel is rendelkezik, de a régi telefon tartalma még mindig velünk van. Hasonlóképpen, egy olyan program, mint a Watson is biztosan megmarad. Ha a Watson hardvere holnap megsemmisülne, a Watson könnyen újra létrehozható lenne a felhőben tárolt biztonsági másolataiból.

Ez egy olyan képességet jelent a nem biológiai világban, amely a biológiai világban nem létezik. Ez előny, nem pedig korlátozás, és ez az egyik oka annak, hogy ma olyan lelkesen töltjük fel emlékeinket a felhőbe. Minden biztonnyal tovább fogunk haladni ebbe az irányba, ahogy a nem biológiai rendszerek egyre inkább elsajátítják biológiai agyunk képességeit.

A dilemma megoldása a következő: Nem igaz, hogy a You 2 nem te vagy - hanem te vagy. Csupán arról van szó, hogy most már ketten vagytok. Ez nem is olyan rossz - ha szerinted te jó dolog vagy, akkor kettő belőled még jobb.

Úgy gondolom, hogy valójában az fog történni, hogy a fokozatos helyettesítés és bővítés forgatókönyvének útján fogunk haladni, míg végül a gondolkodásunk nagy része a felhőben lesz. Az én hitem az identitással kapcsolatban az, hogy az identitás megmarad a minket alkotó információmintázat folytonossága révén. A folyamatosság lehetővé teszi a folyamatos változást, így bár némileg más vagyok, mint tegnap voltam, mégis ugyanazzal az identitással rendelkezem. Az identitásomat alkotó minta folytonossága azonban nem szubsztrátfüggő. A biológiai szubsztrátumok csodálatosak - nagyon messzire vittek minket -, de nagyon jó okokból egy sokkal alkalmasabb és tartósabb szubsztrátot hozunk létre.

FEJEZET 10

A GYORSULÓ MEGTÉRÜLÉS TÖRVÉNYE AZ AGYRA ALKALMAZVA

És ha az ember bizonyos tekintetben meg is marad a magasabb rendű teremtmény, nem felel ez meg a természet gyakorlatának, amely bizonyos dolgokban megengedi a felsőbbrendűséget az állatoknak, amelyeket összességében már régen felülmúltak? Nem engedte-e meg, hogy a hangya és a méh megőrizze felsőbbrendűségét az ember fölött a közösségek szervezésében és a társadalmi berendezkedésben, a madár a levegőben való repülésben, a hal az úszásban, a ló az erőben és a gyorsaságban, a kutya pedig az önfeláldozásban?

-Samuel Butler, 1871

Volt idő, amikor a Föld minden jel szerint teljesen nélkülözte mind az állati, mind a növényi életet, és amikor legjobb filozófusaink véleménye szerint egyszerűen egy forró, kerek gömb volt, amelynek kérge fokozatosan kihűlt. Nos, ha egy emberi lény akkor létezett volna, amikor a Föld ebben az állapotban volt, és úgy láthatta volna, mintha valami más világ lenne, amelyhez semmi köze, és ha ugyanakkor teljesen járatlan lett volna minden fizikai tudományban, vajon nem mondta volna-e ki, hogy lehetetlen, hogy a látszólagos salakból, amelyet látott, tudatossággal rendelkező lények fejlődjenek ki? Nem tagadta volna-e, hogy az a tudatosság bármiféle lehetőségét tartalmazza? Az idő múlásával azonban mégis kialakult a tudat. Nem lehetséges-e tehát, hogy a tudat számára még mindig új csatornák vannak kiasva, bár jelenleg nem észlelhetjük a jeleit?

-Samuel Butler, 1871

Ha elgondolkodunk az élet és a tudatosság már kialakult sokféle fázisán, meggondolatlanság lenne azt állítani, hogy több nem fejlődhet ki, és hogy az állati élet a dolgok vége. Volt idő, amikor a tűz volt minden dolog vége: volt idő, amikor a kőzet és a víz volt az. -Samuel Butler, 1871

A gépi tudatosság végső kialakulása ellen nem jelent biztonságot az a tény, hogy a gépeknek most még kevés tudatosságuk van. Egy puhatestűnek sincs sok tudata. Gondolkodjunk el azon a rendkívüli fejlődésen, amelyet a gépek az elmúlt néhány száz év alatt elértek, és vegyük észre, milyen lassan halad előre az állat- és növényvilág. A magasabban szervezett gépek nem annyira a tegnapi, mint inkább az elmúlt öt perc teremtményei, úgyszólván a múlt időhöz képest. Tegyük fel a vita kedvéért, hogy a tudatos lények mintegy húszmillió éve léteznek: nézzük meg, hogy a gépek az elmúlt ezer évben milyen előrelépéseket tettek! Nem tarthat a világ húszmillió évvel tovább? Ha igen, akkor mivé nem válnak a végén? -Samuel

Az általam gyorsuló hozamok törvényének (LOAR) nevezett tézisének lényege, hogy az információs technológia alapvető mérőszámai kiszámítható és exponenciális pályát követnek, megcáfolva a hagyományos bölcsességet, miszerint "nem lehet megjósolni a jövőt". Még mindig sok dolog van - melyik projekt, vállalat vagy technikai szabvány fog érvényesülni a piacon, mikor lesz béke a Közel-Keleten -, amit nem lehet tudni, de az információ alapját képező ár/teljesítmény és kapacitás mégis figyelemre méltóan kiszámíthatónak bizonyult. Meglepő módon ezeket a tendenciákat nem zavarják az olyan körülmények, mint a háború vagy a béke, a jólét vagy a recesszió.

Az evolúció elsősorban azért hozta létre az agyat, hogy megjósolja a jövőt. Amikor évezredekkel ezelőtt valamelyik ősrünk a szavannákon sétált, észrevehette, hogy egy állat egy olyan útvonal felé tart, amelyen ő is haladt. Megjósolhatta, hogy ha a pályán marad, útjaik keresztezni fogják egymást. Ennek alapján úgy döntött, hogy más irányba indul, és előrelátása értékesnek bizonyult a túlélés szempontjából.

A jövő ilyen beépített előrejelzői azonban lineárisak, nem pedig exponenciálisak, ami a neokortex lineáris szerveződéséből ered. Emlékezzünk, hogy a neokortex folyamatosan jóslatokat készít.

-melyik betűt és szót látjuk legközelebb, kit várunk, ha befordulunk a sarkon, és így tovább. A neokortex lineáris lépéssorozatokkal szerveződik az egyes mintákban, ami azt jelenti, hogy az exponenciális gondolkodás nem természetes számunkra. A kisagy szintén lineáris előrejelzéseket használ. Amikor segít nekünk elkapni egy röptülő labdát, lineáris előrejelzést készít arról, hogy hol lesz a labda a látómezőnkben, és hol kell lennie a kesztyűs kezünknek a látómezőnkben, hogy elkapjuk.

Mint már rámutattam, drámai különbség van a lineáris és az exponenciális fejlődés között (negyven lépés lineárisan negyven, de exponenciálisan egy billió), ami megmagyarázza, hogy a gyorsuló megtérülés törvényéből származó előrejelzéseim miért tűnnek sok megfigyelő számára elsőre meglepőnek. Rá kell szoktatnunk magunkat az exponenciális gondolkodásra. Ha az információs technológiákról van szó, ez a helyes gondolkodásmód.

A gyorsuló megtérülés törvényének kvintesszenciális példája a számítási ár/teljesítmény tökéletesen egyenletes, kétszeresen exponenciális növekedése, amely 110 éven át stabilan tartotta magát két világháború, a nagy gazdasági világválság, a hidegháború, a Szovjetunió összeomlása, Kína újbóli felemelkedése, a közelmúltbeli pénzügyi válság és a 19. század végének, a 20. századnak és a 21. század elejének minden más jelentős eseménye során. Egyesek ezt a jelenséget "Moore törvényének" nevezik, de ez téves felfogás. A Moore-törvény - amely szerint két évente kétszer annyi alkatrészt lehet elhelyezni egy integrált áramkörön, és ezek gyorsabban futnak, mert kisebbek - csak egy paradigma a sok közül. Valójában ez volt az ötödik, nem az első paradigma, amely exponenciális növekedést hozott a számítástechnika ár/teljesítmény arányában.

A számítás exponenciális növekedése az 1890-es amerikai népszámlálással kezdődött (az első automatizált népszámlálással), amely az elektromechanikus számítás első paradigmáját használta, évtizedekkel Gordon Moore születése előtt. A *The Singularity Is*

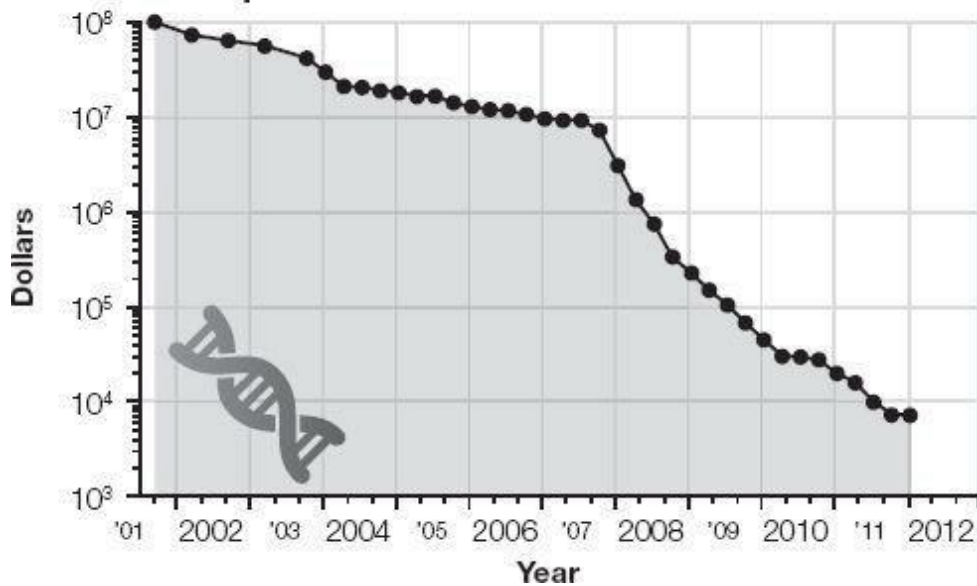
Near (A szingularitás közel van) című könyvben 2002-ig közlöm ezt a grafikont, itt pedig 2009-ig frissítem (lásd a [257.](#) oldalon található grafikont "A számítástechnika exponenciális növekedése 110 éven át" címmel). A simán kiszámítható fejlődési pálya még a közelmúltbeli gazdasági visszaesés idején is folytatódott.

A számítás a legfontosabb példa a gyorsuló megtérülés törvényére, mivel rengeteg adat áll rendelkezésünkre, a számítás mindenütt jelen van, és kulcsszerepet játszik abban, hogy végső soron forradalmasít mindent, amivel foglalkozunk. De messze nem ez az egyetlen példa. Amint egy technológia információs technológiává válik, a LOAR hatálya alá kerül.

A biomedicina az utóbbi időben a technológia és az ipar legjelentősebb, ily módon átalakuló területe. Az orvostudományban a fejlődés történelmileg véletlen felfedezéseken alapult, így a korábbi korszakban a fejlődés lineáris volt, nem pedig exponenciális. Ez mindazonáltal előnyös volt: A várható élettartam az ezer évvel ezelőtti huszonhárom évről kétszáz évvel ezelőtre harminchét évre, napjainkra pedig közel nyolcvan évre nőtt. Az élet szoftverének - a genomnak - az összegyűjtésével az orvostudomány és az emberi biológia információs technológiává vált. Maga az emberi genomprojekt tökéletesen exponenciális volt, a genetikai adatok mennyisége megduplázódott, az egy bázispárra jutó költség pedig minden évben a felére csökkent a projekt kezdete óta 1990.³ (Az ebben a fejezetben szereplő összes grafikon a *The Singularity Is Near (A szingularitás közel van)* megjelenése óta frissült.

Cost per Human Genome

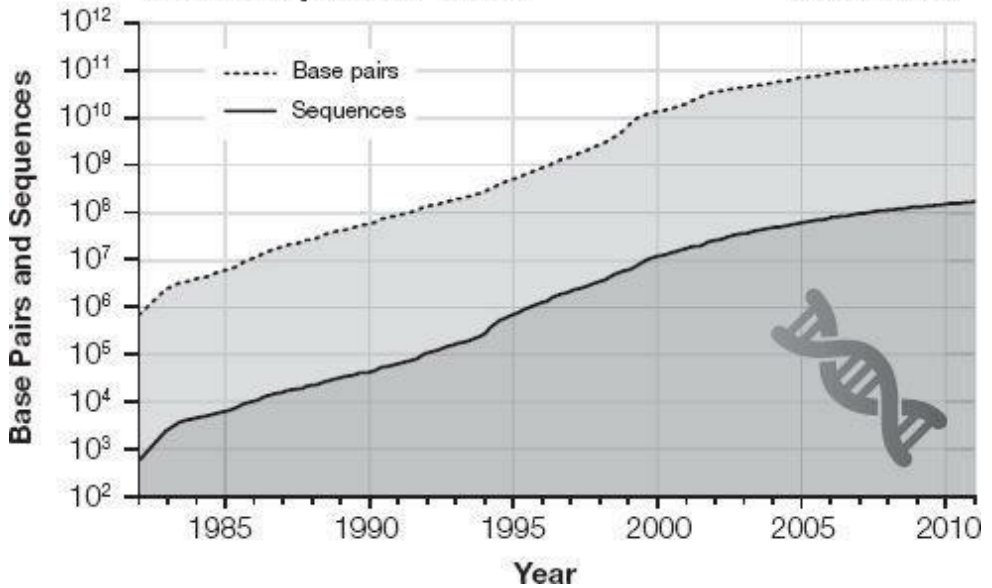
Logarithmic Plot



Egy emberi méretű genom szekvenálásának költségei.¹

Growth in Genbank DNA Sequence Data

Logarithmic Plot



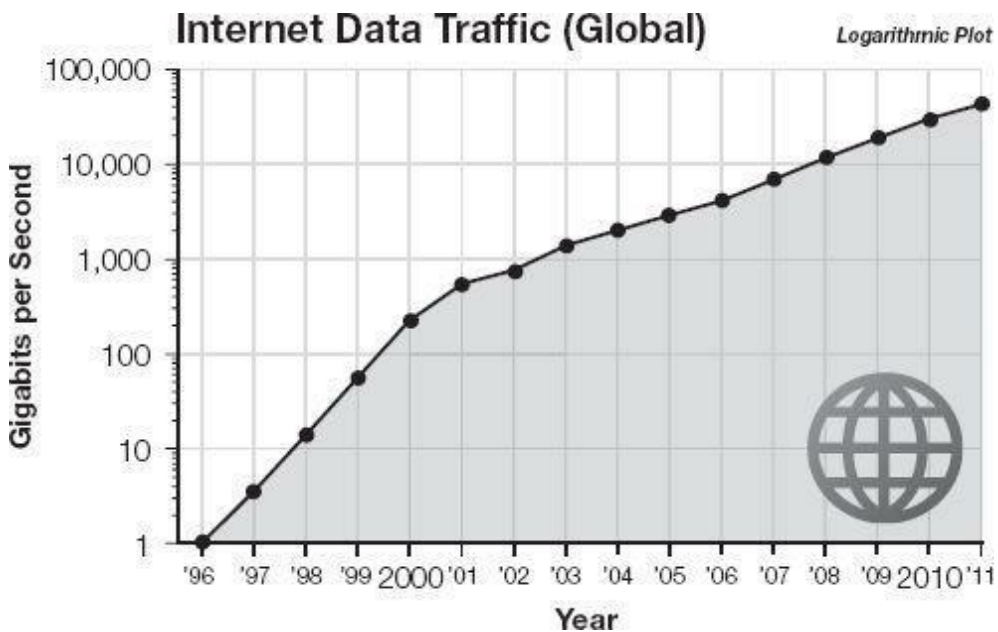
A világon évente szekvenált genetikai adatok mennyisége.²

Ma már képesek vagyunk biomedicinális beavatkozásokat tervezni a számítógépeken és biológiai szimulátorokon tesztelni, amelyek mérete és pontossága szintén évről évre megduplázódik. A saját elavult szoftvereinket is frissíthetjük: Az RNS-interferencia kikapcsolhatja a géneket, a génterápia új formái pedig új géneket adhatnak hozzá, nemcsak egy újszülötthez, hanem egy érett egyedhez is. A genetikai technológiák fejlődése az agy fordított tervezésének projektjét is érinti, mivel ennek egyik fontos aspektusa annak megértése, hogy a gének hogyan irányítják az agyi funkciókat, például új kapcsolatok létrehozását, hogy tükrözzék a nemrég hozzáadott agykérgi tudást. A biológia és az információs technológia integrációjának számos más megnyilvánulása is van, ahogy a genomszekvenáláson túl a genomszintetizálás felé haladunk.

Egy másik információs technológia, amely zökkenőmentes, exponenciális növekedést mutat, az egymással való kommunikáció és az emberi tudás hatalmas tárházainak továbbítása. Ezt a jelenséget sokféleképpen lehet mérni. A Cooper-törvény, amely szerint a vezeték nélküli kommunikáció teljes bitkapacitása egy adott mennyiségű rádióspektrumban harminc havonta megduplázódik, attól kezdve, hogy Guglielmo Marconi 1897-ben a vezeték nélküli távirót Morse-jelek továbbítására használta, egészen a mai 4G kommunikációs technológiáig igaz.⁴ A Cooper-törvény szerint a rádióspektrum adott részén továbbítható információ mennyisége több mint egy évszázada két és fél évente megduplázódik. Egy másik példa erre az interneten másodpercenként továbbított bitek száma, ami

másfél évente megduplázódik.⁵

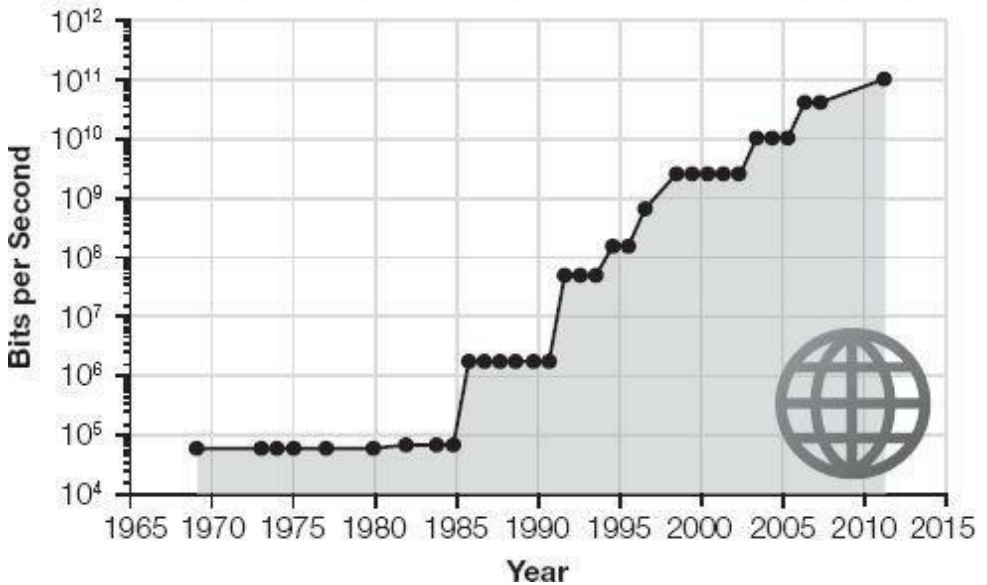
Azért kezdtem el érdeklődni a technológia bizonyos aspektusainak előrejelzése iránt, mert körülbelül harminc évvel ezelőtt rájöttem, hogy a feltalálóként való sikeresség kulcsa (ezt a szakmát öt éves koromban választottam) az időzítésben rejlik. A legtöbb találmány és feltaláló nem azért bukik el, mert maguk a kütyük nem működnek, hanem azért, mert az időzítésük rossz, mert vagy azelőtt jelennek meg, hogy az összes támogató tényező a helyére került volna, vagy túl későn, mert elszalasztották a lehetőség ablakát.



Az internetre szánt nemzetközi (országokénti) sávszélesség a világ számára.⁶

Highest Internet Backbone Bandwidth

Logarithmic Plot



Az internetes gerinchálózat legnagyobb sávszélessége (sebessége).⁷ Mérnökként körülbelül három évtizeddel ezelőtt elkezdtem adatokat gyűjteni a különböző területek technológiai mérőszámairól. Amikor elkezdtem ezt az erőfeszítést, nem számítottam arra, hogy tiszta képet fog mutatni, de reméltem, hogy némi útmutatást ad, és lehetővé teszi számomra, hogy megalapozott találgatásokat tegyek. A célom az volt - és még mindig az -, hogy a saját technológiai erőfeszítéseimet úgy időzítsem, hogy azok megfeleljenek annak a világnak, amely akkor létezik, amikor befejezem a projektet - és amely, mint rájöttem, nagyon különbözni fog attól a világtól, amely akkor létezett, amikor elkezdtem.

Gondoljunk csak bele, hogy a világ mennyit és milyen gyorsan változott a közelmúltban. Néhány évvel ezelőtt az emberek még nem használtak közösségi hálózatokat (a Facebookot például 2012 márciusának végén alapították, és 2004 havi 901 millió aktív felhasználója volt),⁸ wikipédiák, blogok, vagy tweetek. Az 1990-es években a legtöbb ember nem használt keresőmotorokat vagy mobiltelefonokat. Képzeld el a világot ezek nélkül. Ez ókori történelemnek tűnik, pedig nem is volt olyan régen. A közeljövőben a világ még drámaibb változáson megy keresztül.

Nyomozásom során megdöbbentő felfedezést tettem: Ha egy technológia információs technológia, akkor az ár/teljesítmény és a kapacitás (időegységre, költségegységre vagy más erőforrásra vetített) alapvető mérőszámai elképesztően pontos exponenciális pályát követnek.

Ezek a pályák túlszárnyalják a konkrét paradigmákat, amelyeken alapulnak (például a

Moore-törvényt). Amikor azonban egy paradigma kifut a lendületből (például amikor a mérnökök az 1950-es években már nem tudták tovább csökkenteni a vákuumcsövek méretét és költségeit), az kutatási nyomást gyakorol a következő paradigma létrehozására, és így a fejlődés újabb S-görbéje kezdődik.

Az új paradigma következő S-görbéjének exponenciális része aztán folytatja az információtechnológiai intézkedés folyamatos exponenciális részét. Így az 1950-es években a vákuumcsöves számítástechnika az 1960-as években átadta a helyét a tranzistoroknak, majd az 1960-as évek végén az integrált áramköröknek és a Moore-törvénynek, és azon túl. A Moore-törvény viszont utat enged a háromdimenziós számítástechnikának, amelynek korai példái már most is léteznek. Az információs technológiák azért képesek következetesen túllépni bármelyik paradigma korlátain, mert a számításokhoz, a memóriához vagy az információ továbbításához szükséges erőforrások elenyészően kicsik.

Elgondolkozhatunk azon, hogy vannak-e alapvető korlátai annak, hogy képesek vagyunk számításokat végezni és információt továbbítani, függetlenül a paradigmától? A válasz igen, a számítás fizikájáról alkotott jelenlegi ismereteink alapján. Ezek a korlátok azonban nem nagyon korlátozóak. Végso soron a molekuláris számítástechnikán alapuló intelligenciánkat trilliószorosára bővíthetjük. Számításaim szerint még ebben az évszázadban elérjük ezeket a határokat.

Fontos rámutatni, hogy nem minden exponenciális jelenség a gyorsuló hozam törvényének példája. Egyes megfigyelők félreértelmezik a LOAR-t, amikor olyan exponenciális trendekre hivatkoznak, amelyek nem információalapúak: Például rámutatnak, hogy a férfi borotvák egy pengéről kettőről négyre váltak, majd megkérdezik, hol vannak a nyolc pengés borotvák? A borotvák (még) nem számítanak információs technológiának.

A The Singularity Is Near (A szingularitás közel van) című könyvemben elméleti vizsgálatot végzek, beleértve (a könyv függelékében) annak matematikai feldolgozását, hogy a LOAR miért olyan figyelemreméltóan kiszámítható. Lényegében mindig a legújabb technológiát használjuk a következő létrehozásához. A technológiák exponenciális módon épülnek egymásra, és ez a jelenség könnyen mérhető, ha információs technológiáról van szó. Az akkori korszak számítógépeit és egyéb eszközeit 1990használtuk az 1991-es számítógépek létrehozásához; 2012-ben a jelenlegi információs eszközöket használjuk a 2013-as és 2014-es gépek létrehozásához. Tágabb értelemben ez a gyorsulás és exponenciális növekedés minden olyan folyamatra vonatkozik, amelyben az információ mintái fejlődnek. Tehát a biológiai evolúció gyorsulását látjuk, és hasonló (de sokkal gyorsabb) gyorsulást látunk a technológiai evolúcióban, amely maga is a biológiai evolúcióból fejlődött ki.

A gyorsuló megtérülés törvényén alapuló előrejelzéseim már több mint negyed évszázada nyilvánosak, kezdve az 1980-as évek közepén írt *Az intelligens gépek kora című* könyvemben bemutatottakkal. Az abban a könyvben szereplő pontos jóslatok példái közé tartozik: a 90-es évek közepén-végén megjelenik egy hatalmas, világméretű kommunikációs háló, amely összeköti az embereket a világ minden táján egymással és az összes emberi tudással; a demokratizálódás nagy hulláma, amely ebből a decentralizált kommunikációs hálózatból indul ki, és elsöpri a Szovjetuniót; a sakkvilágbajnok legyőzése 1998-ra; és még

sok más.

A gyorsuló hozamok törvényét, ahogyan azt a számításokra alkalmazzák, részletesen leírtam *A szellemi gépek korában*, ahol egy évszázadnyi adatot adtam meg, amely a számítások árának/teljesítményének kétszeresen exponenciális fejlődését mutatja az 1998.alábbiakban2009 frissítettük.

Nemrégiben írtam egy 146 oldalas áttekintést *Az intelligens gépek kora, A spirituális gépek kora és A szingularitás közel van* című könyvekben megfogalmazott előrejelzéseimről. (Az esszét itt olvashatod el, ha az ebben a végjegyzetben található linke mész.)² *A Szellemi gépek kora* több száz előrejelzést tartalmazott konkrét évtizedekre (2009, és2019,2029,

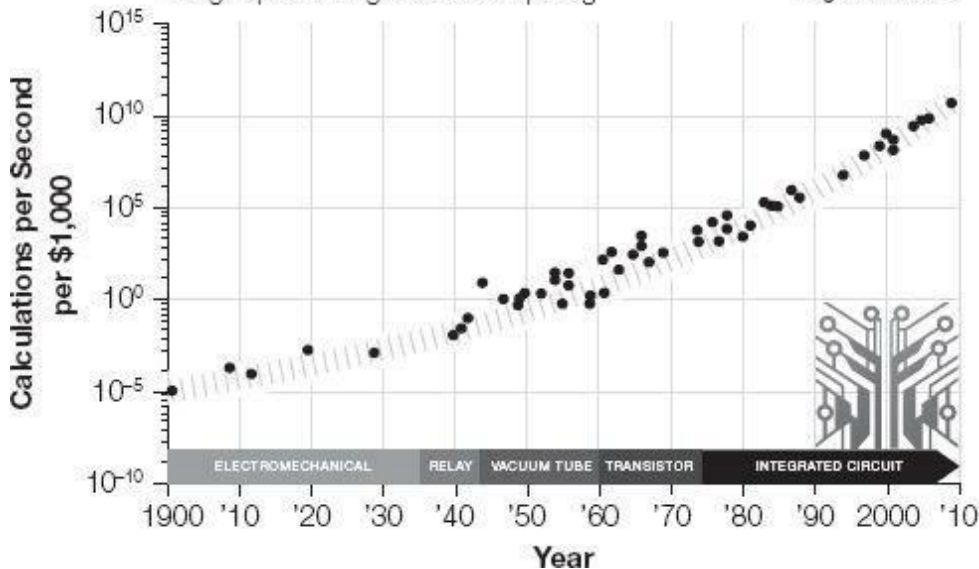
2099). Például *A szellemi gépek kora* című könyvemben, amelyet az 1990-es években írtam, 147 előrejelzést tettem 2009-re. Ezek közül 115 (78 százalék) 2009 végére teljes mértékben helytálló; különösen azok a jóslatok voltak pontosak, amelyek az információs technológiák kapacitásának és ár/teljesítményének alapvető méréseivel foglalkoztak. További 12 (8 százalék) "lényegében helyes". Összesen

127 előrejelzés (86 százalék) helyes vagy lényegében helyes. (Mivel a jóslatok egy adott évtizedre vonatkoztak, egy 2009-re vonatkozó jóslatot akkor tekintettek "lényegében helyesnek", ha az 2010-ben vagy 2011-ben vált valóra.) További 17 (12 százalék) részben helyes, 3 (2 százalék) pedig téves.

Exponential Growth of Computing for 110 Years

Moore's law was the fifth, not the first, paradigm to bring exponential growth in computing

Logarithmic Plot

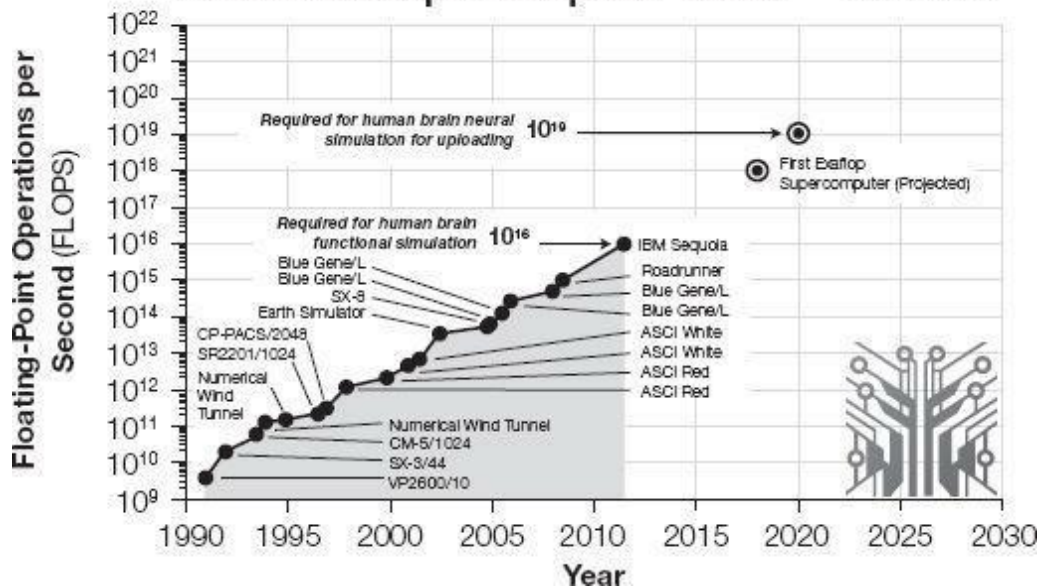


A különböző számítástechnikai eszközök másodpercenkénti számításai (állandó)

ezer dollárra vetítve.¹⁰

Growth in Supercomputer Power

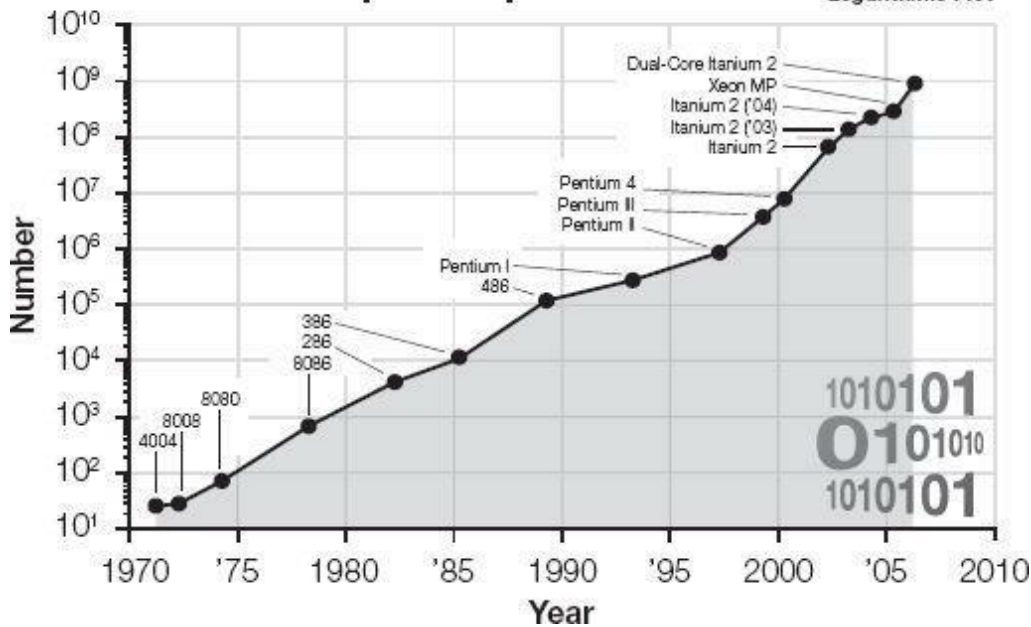
Logarithmic Plot



Lebegőpontos műveletek per másodperc of különböző szuperszámítógépek. [1](#)

Transistors per Chip

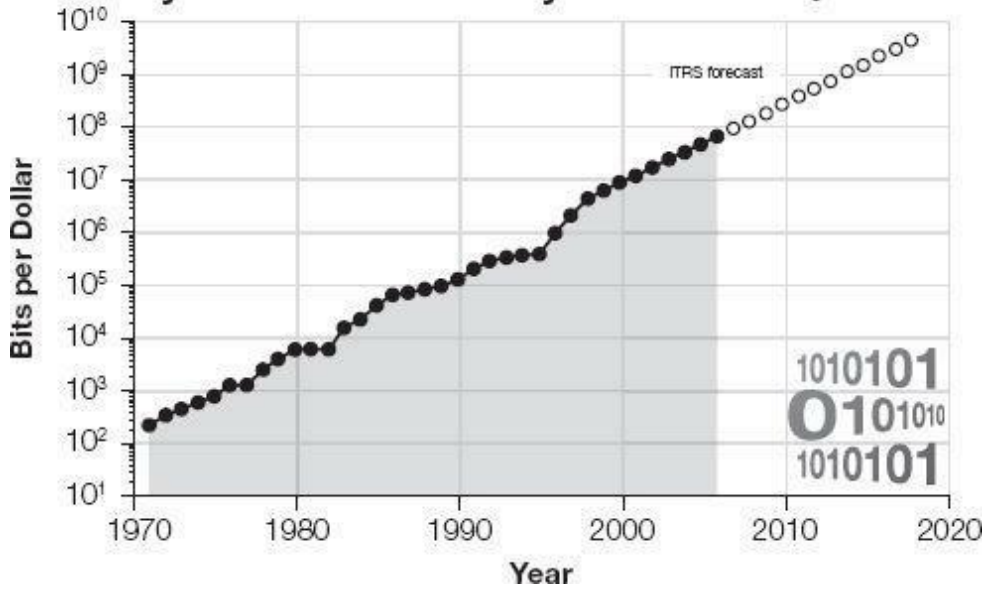
Logarithmic Plot



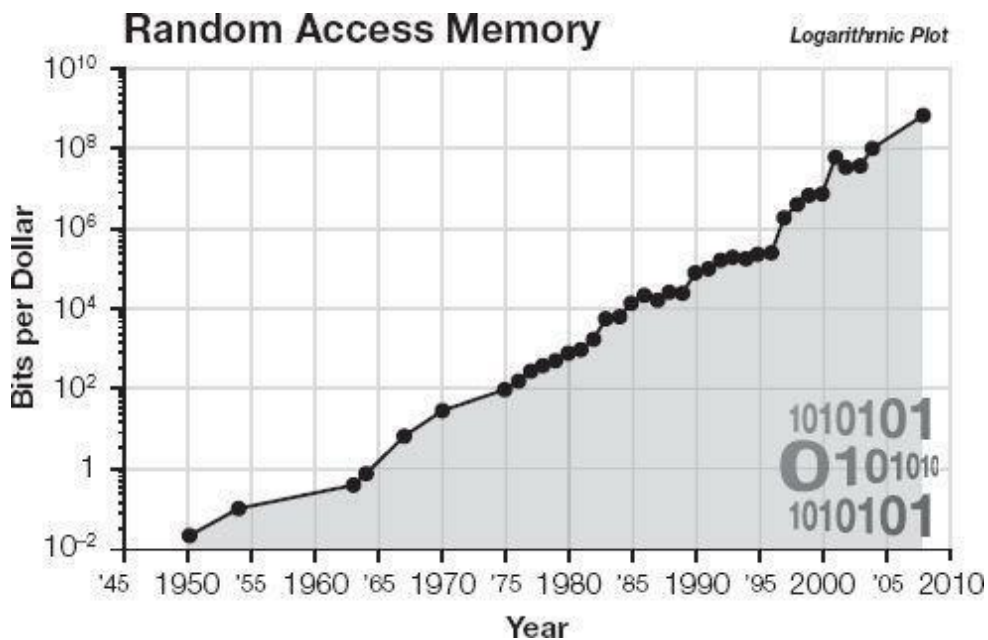
Tranzisztorok chipenként a különböző Intel processzorok esetében.¹²

Dynamic RAM Memory

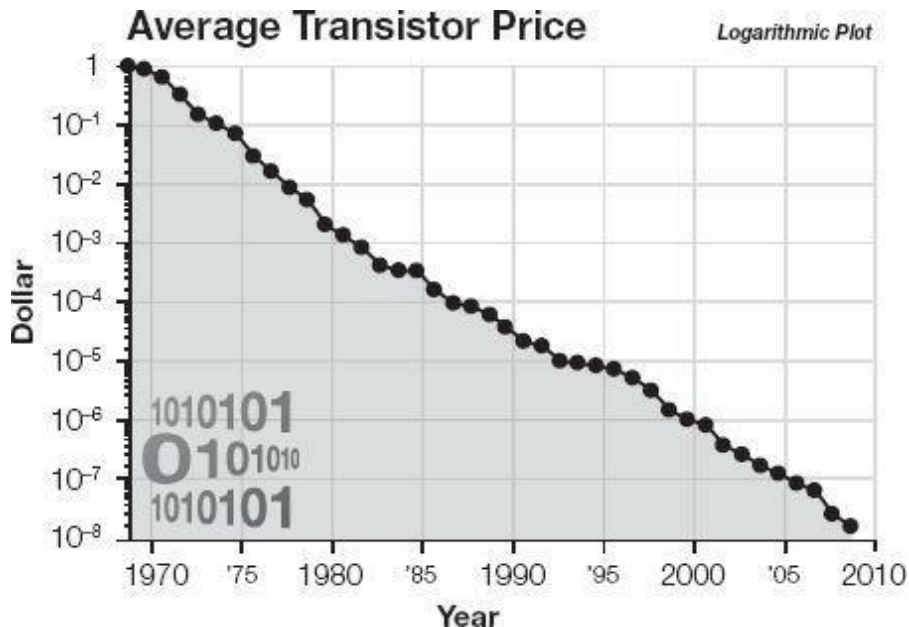
Logarithmic Plot



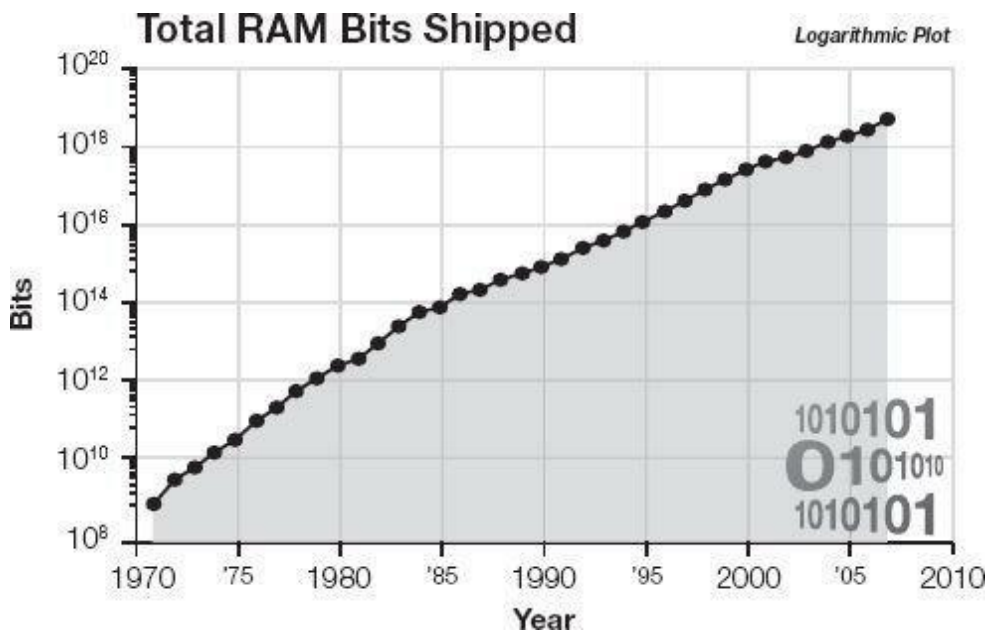
A dinamikus véletlen hozzáférésű memóriachipek dolláronkénti bitjei.¹³



Bits per dollár a véletlen hozzáférésű memóriachipek esetében.¹⁴

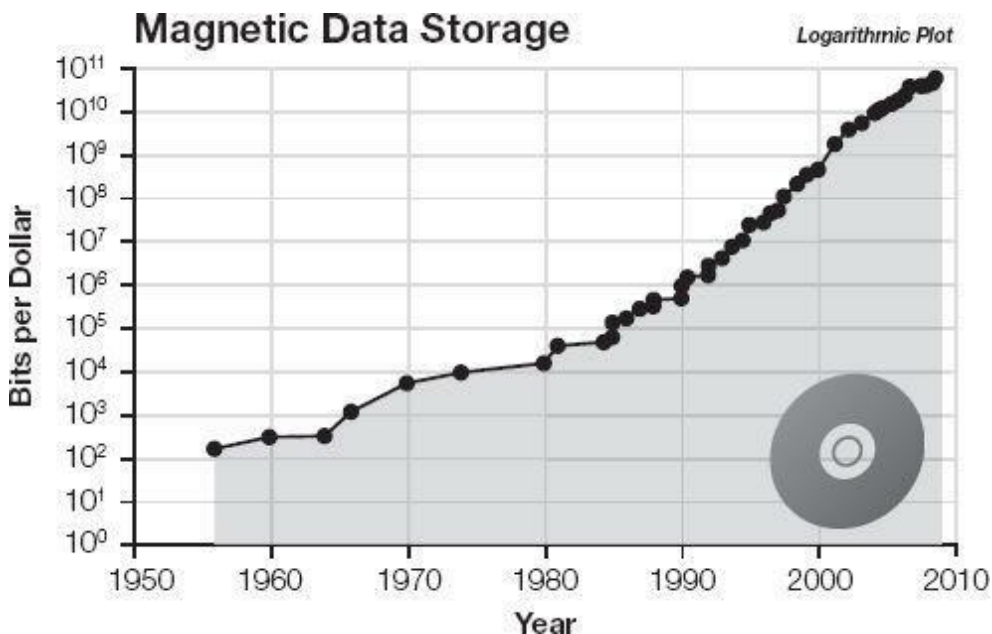


Az átlagos tranzisztoronkénti ár dollárban.¹⁵



A véletlen hozzáférésű memória szállított bitjeinek teljes száma

minden évben.¹⁶



Bitek dolláronként (2000-es állandó dollárban) mágneses adattárolás esetén.¹⁷

Még a "téves" előrejelzések sem voltak mind tévesek.

Például tévesnek ítélem meg azt a jóslatot, hogy önvezető autókat fogunk kapni, annak ellenére, hogy a Google bemutatta az önvezető autókat, és annak ellenére, hogy 2010 októberében négy vezető nélküli elektromos furgon sikeresen teljesített egy 13 000 kilométeres tesztutat Olaszországtól Olaszorszáig.

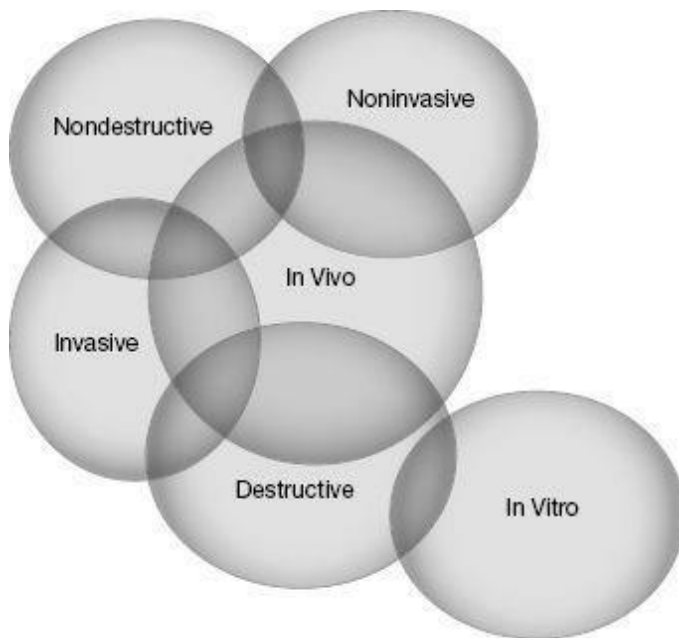
Kína.¹⁸ A terület szakértői jelenleg azt jósolják, hogy ezek a technológiák az évtized végére már rutinszerűen elérhetővé válnak a fogyasztók számára.

Az exponenciálisan bővülő számítástechnikai és kommunikációs technológiák mind hozzájárulnak az emberi agy módszereinek megértésére és újratereztetésére irányuló projekthez. Ez a törekvés nem egyetlen szervezett projekt, hanem inkább sokféle projekt eredménye, beleértve az agy összetevőinek részletes modellezését az alábbiaktól kezdve a következő területeken

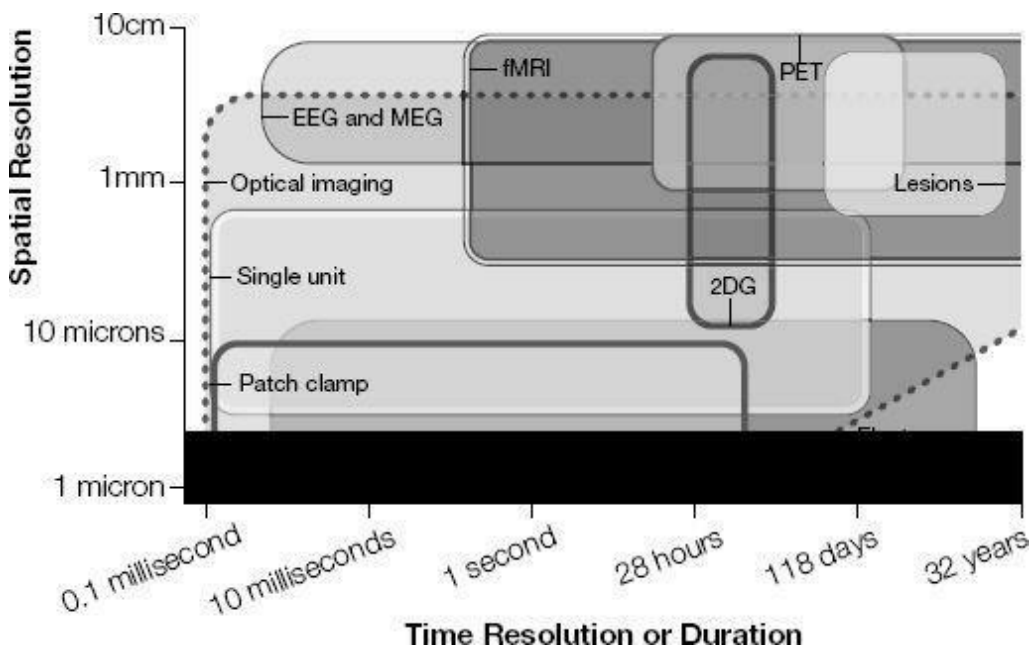
az egyes neuronoktól a teljes neokortexig, a "connectome" (az agyban lévő idegi kapcsolatok) feltérképezése, az agyi régiók szimulációja és még sok más. Ezek mindegyike exponenciálisan növekszik. A könyvben bemutatott bizonyítékok nagy része csak a közelmúltban vált elérhetővé - például a [fejezetben](#) tárgyalt 2012-es Wedeen-tanulmány, amely [4](#) kimutatta a neokortexben lévő kapcsolatok nagyon rendezett és "egyszerű" (a kutatókat idézve) rácsszerű mintázatát. Az említett tanulmány kutatói elismerik, hogy felismerésük (és képeik) csak az új, nagy felbontású képalkotó technológia eredményeként váltak megvalósíthatóvá.

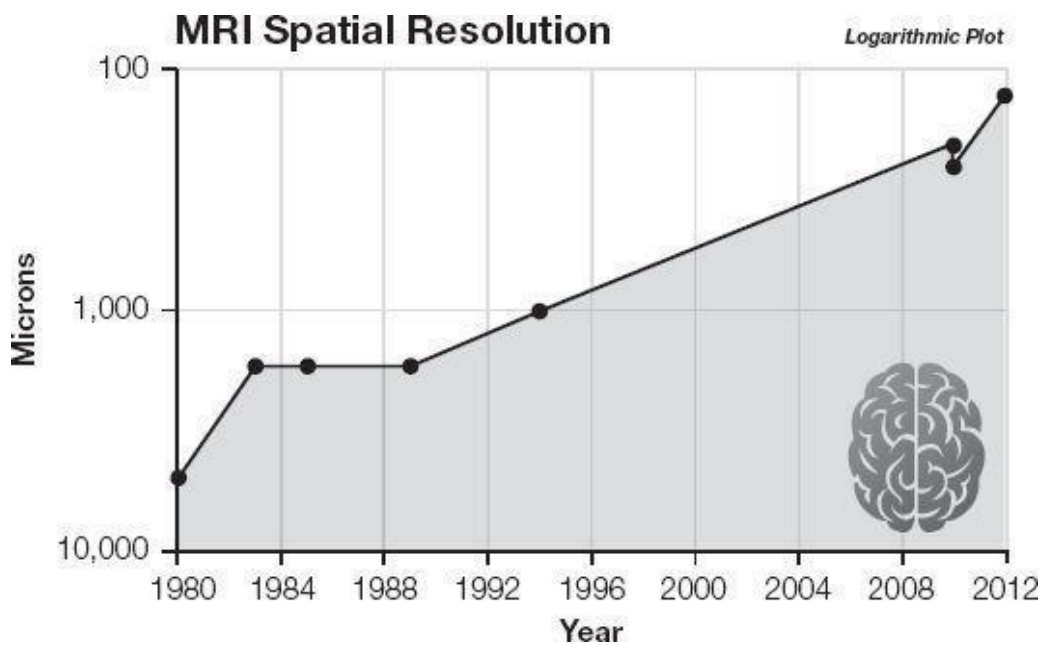
Az agyszkenelési technológiák térbeli és időbeli felbontása exponenciális ütemben javul. Az agyszkenelési módszerek különböző típusai a teljesen noninvazív, embereken alkalmazható módszerektől az állatokon alkalmazott invazívabb vagy pusztítóbb módszerekig terjednek.

Az MRI (mágneses rezonancia képalkotás), egy viszonylag nagy időbeli felbontású, nem invazív képalkotó technika, folyamatosan, exponenciális ütemben fejlődött, és mára a térbeli felbontás megközelíti a 100 mikrométert (a méter milliomod részét).



Az agyi képalkotó módszerek Venn-diagramja.¹⁹

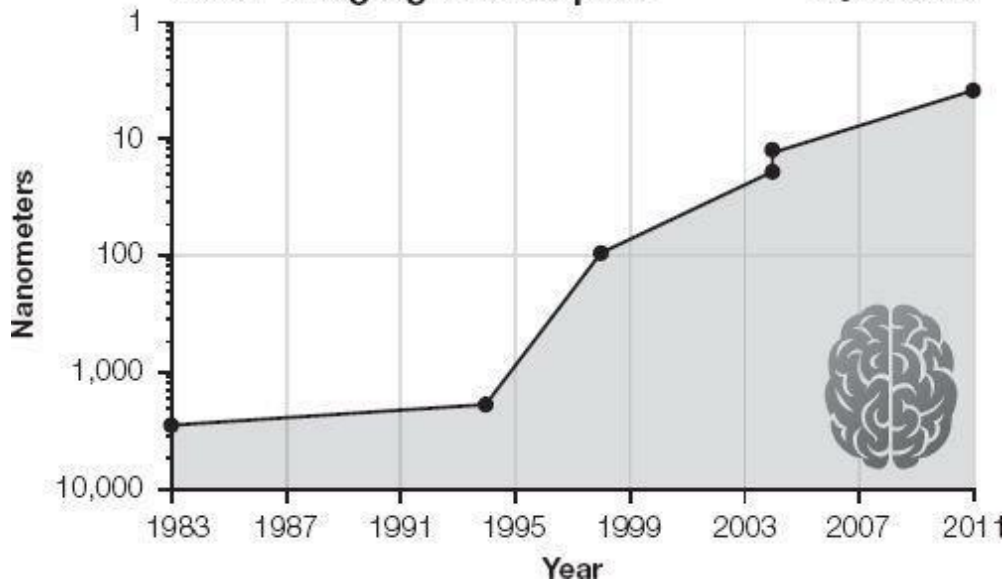




MRI térbeli felbontás mikronokban.²¹

Spatial Resolution of Destructive Brain Imaging Techniques

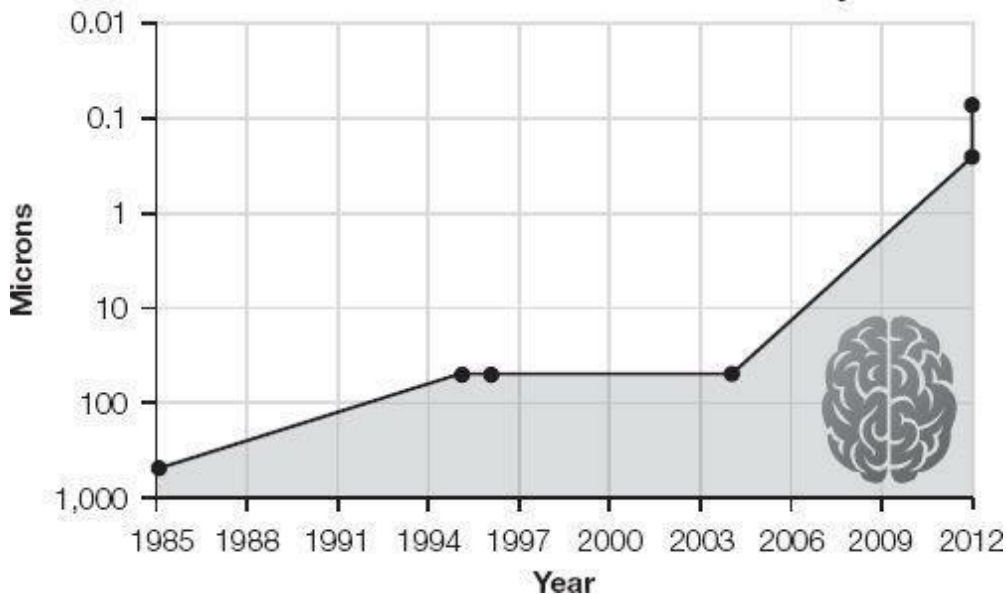
Logarithmic Plot



A roncsolásos képalkotó eljárások térbeli felbontása.²²

Nondestructive Brain Imaging Resolution in Animals

Logarithmic Plot



A roncsolásmentes képalkotó technikák térbeli felbontása állatokban.²³

Romboló képalkotás, amelyet azért végeznek, hogy összegyűjtsék a

connectome (az összes interneuronális kapcsolat térképe) az állati agyakban szintén exponenciális ütemben fejlődött. A jelenlegi maximális felbontás körülbelül négy nanométer, ami elegendő az egyes kapcsolatok megismeréséhez.

Az olyan mesterséges intelligencia-technológiákat, mint a természetes nyelvet megértő rendszerek, nem feltétlenül az agyműködés elméleti alapelveinek utánzására, hanem inkább a maximális hatékonyságra tervezték. Ennek fényében figyelemre méltó, hogy a győztes technikák összhangban vannak az általam ebben a könyvben felvázolt elvekkkel: az invariáns önasszociatív minták önszerveződő, hierarchikus felismerői, redundanciával és felle előrejelzésekkel. Ezek a rendszerek exponenciálisan skálázódnak, amint azt Watson is bizonyította.

Az agy megértésének elsődleges célja az, hogy bővítsük a

intelligens rendszerek létrehozásához szükséges technikák eszköztára. Bár sok mesterséges intelligencia-kutató talán nem teljesen érti ezt, de az agy működésének alapelveiről szerzett ismereteink már most is nagy hatással vannak rájuk. Az agy megértése abban is segít, hogy visszafordítsuk a különböző agyi működési zavarokat. Az agy visszafejtésére irányuló projektnek természetesen van egy másik kulcsfontosságú célja is: annak megértése, hogy kik vagyunk.

FEJEZET 11 TEKINTETEK

Ha egy gépről bebizonyosodik, hogy megkülönböztethetetlen az embertől, akkor ugyanolyan tiszteletet kell tanúsítanunk iránta, mint amelyet az ember iránt tanúsítanánk - el kell fogadnunk, hogy van elméje.

-Stevan Harnad

A gyorsuló megtérülés törvényéről szóló tézissel és annak az emberi intelligencia felerősítésére való alkalmazásával szembeni legjelentősebb ellenvetés az emberi intuíció lineáris természetéből ered. Amint azt korábban leírtam, a neokortexben található több százmillió mintafelismerő mindegyike szekvenciálisan dolgozza fel az információt. Ennek a szerveződésnek az egyik következménye, hogy lineáris elvárásaink vannak a jövővel kapcsolatban, ezért a kritikusok lineáris intuíciójukat olyan információs jelenségekre alkalmazzák, amelyek alapvetően exponenciálisak.

Az ilyen irányú ellenvetéseket "hitetlenségből fakadó kritikának" nevezem, mivel az exponenciális előrejelzések hihetetlennek tűnnek lineáris előszeretetünk miatt, és sokféle formát ölthetnek. A Microsoft társalapítója, Paul Allen (1953-ban született) és kollégája, Mark Greaves nemrégiben több ilyen is megfogalmazott a "The Singularity Isn't" című esszéjében.

Közel" címmel jelent meg a *Technology Review* magazinban.¹ Bár itt Allen konkrét kritikáira reagálok, ezek a kritikák az általam felhozott érvekkel szembeni ellenvetések tipikus körét képviselik, különösen az agyra vonatkozóan. Bár Allen a *The Singularity Is Near (A szingularitás közel van)* című esszéjének címében hivatkozik, az egyetlen idézet a cikkben egy 2001-ben írt esszéjére ("The Law of Accelerating Returns"). Ráadásul a cikke nem ismeri el és nem válaszol azokra az érvekre, amelyeket a könyvben kifejtet. Sajnos gyakran tapasztalom, hogy ez a helyzet a munkámat bírálókkal.

Amikor *A szellemi gépek kora* 1999-ben megjelent, később az esszével²⁰⁰¹ kiegészítve, több kritikát is kiváltott, például a következőket: *Az agy túl bonyolult; az agyban vannak olyan képességek, amelyek természetüknél fogva nem reprodukálhatók szoftverben; és*

még számos más. Az egyik ok, amiért megírtam a *The Singularity Is Near című könyvet*, az volt, hogy válaszoljak ezekre a kritikákra.

Nem mondhatom, hogy Allent és a hasonló kritikusokat feltétlenül meggyőzték volna a könyvben kifejtett érveim, de legalább ő és mások reagálhattak volna arra, amit valójában írtam. Allen azt állítja, hogy "a gyorsuló megtérülés törvénye (LOAR)... nem fizikai törvény". Rámutatnék arra, hogy a legtöbb tudományos törvény nem fizikai törvény, hanem alacsonyabb szinten nagyszámú esemény emergens tulajdonságaiból adódik. Klasszikus példa erre a termodinamikai törvények (LOT). Ha megnézzük a LOT alapjául szolgáló matematikát, az minden egyes részecskét úgy modellez, mintha véletlen sétát követne, így definíció szerint nem tudjuk megjósolni, hogy egy adott részecske hol lesz egy jövőbeli időpontban. A gáz általános tulajdonságai azonban a termodinamika *törvényei szerint* nagy pontossággal megjósolhatóak. Így van ez a gyorsuló visszatérés törvényével is: Minden egyes technológiai projekt és hozzájáruló kiszámíthatatlan, de az általános pálya - az ár/teljesítmény és a kapacitás alapvető mérőszámaival számszerűsítve - mégis figyelemre méltóan kiszámítható pályát követ.

Ha a számítástechnikát csak egy maroknyi kutató folytatná, az valóban kiszámíthatatlan lenne. De ez a technológia a versengő projektek kellően dinamikus rendszerének terméke, hogy az ár/teljesítmény alapvető mérőszáma, mint például a másodpercenkénti számítások állandó dollárra vetítve, egy nagyon egyenes exponenciális pályát követ, amely az amerikai 1890népszámlálásig nyúlik vissza, amint azt az [előző fejezetben](#) megjegyeztem. Bár a LOAR elméleti alapjait részletesen bemutatja a *The Singularity Is Near (A szingularitás közel van) című* könyv, a legerősebb érv mellette az általam és mások által bemutatott kiterjedt empirikus bizonyítékok.

Allen azt írja, hogy "ezek a 'törvények' addig működnek, amíg nem működnek". Itt összekeveri a paradigmákat az informatika egy alapvető területének folyamatos fejlődési pályájával. Ha például az egyre kisebb vákuumcsövek létrehozásának tendenciáját vizsgálánk - a számítások javításának paradigmáját az 1950-es években -, akkor igaz, hogy ez addig folytatódott, amíg nem folytatódott. De ahogy ennek a paradigmának a vége világossá vált, a kutatási nyomás egyre nőtt a következő paradigma felé. A tranzisztorok technológiája fenntartotta a számítások ár/teljesítmény exponenciális növekedésének alapvető trendjét, és ez vezetett az ötödik paradigmához (Moore törvénye) és az integrált áramkörök jellemzőinek folyamatos tömörítéséhez. Rendszeresen hangzottak el olyan előrejelzések, hogy a Moore-törvény véget fog érni. A félvezetőipar "A félvezetőkre vonatkozó nemzetközi technológiai ütemterv" a következőket tervezi

hét nanométeres jellemzőkkel a 2020-as évek elejére.² Ekkor a kulcsfontosságú jellemzők harmincöt szénatom szélességűek lesznek, és nehéz lesz tovább zsugorítani őket. Az Intel és más chipgyártók azonban már most megteszik az első lépéseket a hatodik paradigma, a három dimenzióban történő számítás felé, hogy az ár/teljesítmény exponenciális javulása folytatódjon. Az Intel előrejelzése szerint a háromdimenziós chippek a tizenévesekre válnak általánossá; háromdimenziós tranzisztorokat és háromdimenziós memóriachipeket már bevezettek. Ez a hatodik paradigma fogja fenntartani a LOAR-t a számítógépek ár/teljesítménye tekintetében egy olyan időpontig, még ebben a században, amikor az ezer dollár értékű számítási kapacitás trilliószor erősebb lesz, mint az emberi agy.³ (Úgy tűnik, hogy Allen és én legalábbis egyetértünk abban, hogy milyen szintű számítás szükséges az

emberi agy funkcionális szimulációjához).⁴ Allen ezután a szokásos érveléssel folytatja, miszerint a szoftverek nem fejlődnek ugyanolyan exponenciális ütemben, mint a hardverek. *A The Singularity Is Near (A szingularitás közel van)* című könyvemben hosszasan foglalkoztam ezzel a kérdéssel, idézve a szoftverek komplexitásának és képességének mérésére szolgáló különböző módszereket, amelyek hasonló exponenciális növekedést mutatnak.⁵ Egy nemrégiben készült tanulmány ("Jelentés az elnöknek és a kongresszusnak, A digitális jövő tervezése: Az Elnök Tudományos és Technológiai Tanácsadói Tanácsa) a következőket állapítja meg:

Még figyelemreméltóbb - és még kevésbé közismert -, hogy sok területen az *algoritmusk* *javulásának köszönhető teljesítménynövekedés még a processzorok sebességének növekedése miatti drámai teljesítménynövekedést is messze meghaladta.* Az algoritmusok, amelyeket ma a beszédfelismeréshez, a természetes nyelvi fordításhoz, a sakkozáshoz, a logisztikai tervezéshez használunk, az elmúlt évtizedben figyelemre méltóan fejlődtek. Íme csak egy példa, amelyet Martin Grötschel professzor, a berlini Konrad-Zuse-Zentrum für Informationstechnik professzora adott. Grötschel, aki az optimalizálás szakértője, megjegyzi, hogy egy lineáris programozással megoldott referencia termelési tervezési modell megoldása 1988-ban 82 évig tartott volna az akkori számítógépek és lineáris programozási algoritmusok segítségével. Tizenöt évvel később - 2003-ban - ugyanezt a modellt nagyjából 1 perc alatt lehetett megoldani, ami nagyjából 43 milliószoros javulást jelent. Ebből nagyjából 1000-szeresét a processzorok sebességének növekedése, míg nagyjából kétszeresét az algoritmusok fejlődése 43,000-ozta! Grötschel a 30,000-vegyes egészértékű programozás esetében is nagyjából algoritmikus javulást említ a és 1991 között Az algoritmusok 2008.tervezése és elemzése, valamint a problémák eredendő számítási bonyolultságának tanulmányozása a számítástechnika alapvető részterületei.

Megjegyzendő, hogy a lineáris programozás, amelyet Grötschel fentebb idéz, és amely a teljesítmény milliószoros⁴³ javulását eredményezte, az a matematikai technika, amelyet az erőforrások optimális kiosztására használnak egy olyan hierarchikus memóriarendszerben, mint a HHMM, amelyet korábban tárgyaltam. Sok más hasonló példát is idézek a *The*

*A szingularitás közel van.*⁶

Ami a mesterséges intelligenciát illeti, Allen gyorsan elutasítja az IBM Watsonját, és ezt a véleményt sok más kritikus is osztja. E kritikusok közül sokan semmit sem tudnak a Watsonról azon kívül, hogy az egy számítógépen futó szoftver (bár párhuzamos, 720 processzormaggal). Allen azt írja, hogy a Watsonhoz hasonló rendszerek "törékenyek maradnak, teljesítményük határait mereven a belső feltételezéseik és a meghatározó algoritmusaik szabják meg, nem tudnak általánosítani, és gyakran adnak értelmetlen válaszokat a saját területükön kívül".

Először is, hasonló megfigyelést tehetnénk az emberekről is. Arra is rámutatnék, hogy Watson "speciális területei" magukban foglalják a Wikipédia *egészét*, valamint számos más tudásbázist, ami aligha jelent szűk fókuszot. Watson az emberi tudás széles skálájával foglalkozik, és képes a nyelv finom formáinak kezelésére, beleértve a szójátékokat,

haszonlatokat és metaforákat az emberi törekvések gyakorlatilag minden területén. Nem tökéletes, de az emberek sem azok, és elég jó volt ahhoz, hogy a *Jeopardy!* versenyen a legjobb emberi játékosok felett is győzedelmeskedjen.

Allen azt állítja, hogy Watson maguk a tudósok állították össze, a szűk körű ismeretek minden egyes láncszemét sajátos területeken építve. Ez egyszerűen nem igaz. Bár Watson adatainak néhány területét közvetlenül programozták, tudásának jelentős többségét Watson magától szerezte meg, természetes nyelvű dokumentumok, például a Wikipédia olvasásával. Ez jelenti a legfontosabb erősségét, ahogyan az is, hogy képes megérteni a *Jeopardy!* lekérdezések (kérdésre keresett válaszok) szövevényes nyelvezetét.

Ahogy korábban említettem, a Watsonra vonatkozó kritikák nagy része az, hogy a statisztikai valószínűségeken keresztül dolgozik, nem pedig a "valódi" megértésen keresztül. Sok olvasó ezt úgy értelmezi, hogy Watson csupán statisztikákat gyűjt a szósorozatokról. A "statisztikai információ" kifejezés Watson esetében valójában az olyan önszerveződő módszerekben, mint például a hierarchikus rejtett Markov-modellekben használt elosztott együtthatókra és szimbolikus kapcsolatokra utal. Ugyanilyen könnyen el lehetne utasítani az emberi agykéregben lévő elosztott neurotranszmitter-koncentrációkat és redundáns kapcsolati mintákat is "statisztikai információnak". Valójában ugyanúgy oldjuk fel a kétértelműségeket, ahogy Watson teszi - figyelembe véve egy mondat különböző értelmezéseinek valószínűségét.

Allen így folytatja: "Minden szerkezetet [az agyban] az evolúció évmilliói során pontosan úgy alakítottak ki, hogy egy bizonyos dolgot csináljon, bármi legyen is az. Ez nem olyan, mint egy számítógép, amelyben milliárdnyi egyforma tranzisztor van szabályos memóriamezőkben, amelyeket egy néhány különböző elemmel rendelkező CPU vezérel. Az agyban minden egyes struktúrát és idegi áramkört egyedileg finomított az evolúció és a környezeti tényezők."

Az az állítás, hogy az agyban minden struktúra és idegpálya egyedi és eleve ott van, egyszerűen lehetetlen, mert ez azt jelentené, hogy az agy tervrajzához több száz trillió bájtnyi információra lenne szükség. Az agy szerkezeti tervét (akárcsak a test többi részének tervét) a genom tartalmazza, és maga az agy nem tartalmazhat több tervezési információt, mint a genom. Megjegyzendő, hogy az epigenetikai információ (például a génexpressziót szabályozó peptidok) nem növeli számottevően a genomban lévő információ mennyiségét. A tapasztalat és a tanulás jelentősen növeli az agyban található információ mennyiségét, de ugyanez elmondható az olyan mesterséges intelligencia rendszerekről is, mint a Watson. A *The Singularity Is Near* című könyvemben bemutatom, hogy a veszteségmentes tömörítés után (a genomban lévő masszív redundancia miatt) a genomban lévő tervezési információ mennyisége körülbelül 50 millió bájt, amelynek nagyjából a fele (azaz körülbelül 25 millió bájt).

az agyra vonatkozik.⁷ Ez nem egyszerű, de a komplexitásnak egy olyan szintje, amellyel meg tudunk birkózni, és kisebb komplexitást jelent, mint a modern világ számos szoftverrendszere. Ráadásul az agy 25 millió bájtnyi genetikai tervezési információjának nagy része a neuronok biológiai követelményeire vonatkozik, nem pedig az információfeldolgozó algoritmusokra.

Hogyan jutunk el az agyban 100-1000 trilliónyi kapcsolathoz mindössze több tízmillió bájtnyi tervezési információból? A válasz nyilvánvalóan a masszív redundancia.

Dharmendra Modha, az IBM Research kognitív számítástechnikai részlegének vezetője azt írja, hogy "a neuroanatómusok nem egy reménytelenül kusza, tetszőlegesen összekapcsolt hálózatot találtak, amely teljesen sajátos az egyes egyedek agyában, hanem ehelyett nagyszámú ismétlődő struktúrát az egyes agyakon belül, és nagyszámú homológiát a fajok között..... A meglepő természetes rekonfigurálhatóság reményt ad arra, hogy a neuroszámítás alapvető algoritmusai függetlenek a konkrét érzékszervi vagy motoros modalitásoktól, és hogy az agykérgi struktúrában megfigyelt, területek közötti eltérések nagy része egy kanonikus áramkör finomítását jelenti; ez valóban ez a kanonikus áramkör.

amit vissza akarunk fejleszteni."⁸

Allen mellett érvel, hogy az emberi agy megértésében és képességeinek megisméltésében szükségszerűen korlátozza a fejlődést a benne rejlő "komplexitásfék", amely azon a felfogásán alapul, hogy az emberi agyban lévő körülbelül 100trilliónyi 1,000 kapcsolat mindegyike

ott kifejezett szándékkal. Az ő "komplexitásfékje" összekeveri az erdőt a fákkal. Ha meg akarunk érteni, modellezni, szimulálni és újra létrehozni egy hasnyálmirigyét, akkor nem kell minden egyes hasnyálmirigy-szigetsejt minden egyes organelláját újra létrehozni vagy szimulálni. Ehelyett inkább egy szigetsejtet szeretnénk megérteni, majd absztrahálni annak alapvető működését az inzulinszabályozás szempontjából, majd ezt kiterjeszteni az ilyen sejtek nagy csoportjára. Ez az algoritmus jól érthető a szigetsejtek tekintetében. Jelenleg olyan mesterséges hasnyálmirigyek vannak tesztelés alatt, amelyek ezt a funkcionális modellt használják. Bár az agyban minden bizonnyal sokkal bonyolultabb és változatosabb a működés, mint a hasnyálmirigy tömegesen ismétlődő szigetsejtjeiben, ennek ellenére a funkcióknak mégis masszív ismétlődése van, ahogyan azt ebben a könyvben többször is leírtam.

Az Allen-féle kritikák azt is megfogalmazzák, amit én "tudós pesszimizmusnak" nevezek. A kutatók, akik egy technológia következő generációján vagy egy tudományterület modellezésén dolgoznak, kivétel nélkül az adott közvetlen kihívásokkal küzdenek, így ha valaki leírja, hogy a technológia hogyan fog kinézni tíz generáció múlva, a szemük elkerekedik. Az integrált áramkörök egyik úttörője nemrégiben felidézte nekem, hogy harminc évvel ezelőtt milyen nehézségek árán jutottak el a 10 mikronos (10 000 nanométer) jellemzőméretről az 5 mikronos (5 000 nanométer) jellemzőkig. A tudósok óvatosan bíztak abban, hogy eléri ezt a célt, de amikor az emberek megjósolták, hogy egy nap valóban 1 mikron (1000 nanométer) alatti jellemzőmérettel rendelkező áramkörökkel fogunk rendelkezni, a legtöbbjük, a saját céljukra koncentrálna, úgy gondolta, hogy ez túl vad gondolat ahhoz, hogy belegondoljanak. Ellenvetések hangzottak el az áramkörök törekenységére, a hőhatásokra és így tovább. Ma az Intel már 22 nanométeres kapuhosszúságú chipeket kezd használni.

Ugyanilyen pesszimizmusnak lehetünk tanúi a humán genomprojekt kapcsán is. A tizenöt éves erőfeszítés felénél csak a 1 százalékát gyűjtötték össze a genomnak, és a kritikusok alapvető korlátokat javasoltak arra vonatkozóan, hogy milyen gyorsan lehet szekvenálni anélkül, hogy a kényes genetikai struktúrákat tönkretennék. De a kapacitás és az ár/teljesítmény exponenciális növekedésének köszönhetően a projekt hét évvel később befejeződött. Az emberi agy visszafejtésére irányuló projekt is hasonlóan halad. Csak nemrégiben értünk el például egy olyan küszöböt a nem invazív szkennelési technikákkal, hogy valós időben láthatjuk az egyes interneuronális kapcsolatok kialakulását és tüzelését. Az ebben a könyvben bemutatott bizonyítékok nagy része ilyen fejlesztéseken múlt, és csak nemrég vált elérhetővé.

Allen úgy írja le az emberi agy visszafejtésére vonatkozó javaslatomat, mintha egyszerűen csak átvizsgálnánk az agyat, hogy megértsük annak finom szerkezetét, majd "alulról felfelé" szimulálnánk egy teljes agyat anélkül, hogy megértenénk annak információfeldolgozási módszereit. Ez nem az én javaslatom. Részletesen meg kell értenünk, hogyan működnek az egyes neurontípusok, majd információt kell gyűjtenünk arról, hogy a funkcionális modulok hogyan kapcsolódnak egymáshoz. Az ilyen típusú elemzésből származó funkcionális módszerek aztán irányíthatják az intelligens rendszerek fejlesztését. Alapvetően olyan biológiailag inspirált módszereket keresünk, amelyek

felgyorsíthatják a mesterséges intelligencia területén végzett munkát, amelynek nagy része anélkül haladt előre, hogy jelentős betekintést nyert volna abba, hogy az agy hogyan látja el a hasonló funkciókat. A beszédfelismeréssel kapcsolatos saját munkámból tudom, hogy a munkánk nagymértékben felgyorsult, amikor betekintést nyertünk abba, hogy az agy hogyan készíti elő és alakítja át a hallási információkat.

Az agy tömegesen redundáns struktúrái a tanulás és a tapasztalat révén differenciálódnak. A mesterséges intelligencia jelenlegi fejlettségi szintje valóban lehetővé teszi a rendszerek számára, hogy saját tapasztalataikból is tanuljanak. A Google önvezető autói a saját vezetési tapasztalataikból, valamint az emberi sofőrök által vezetett Google-autók adataiból tanulnak; Watson a tudása nagy részét saját magától tanulta meg. Érdekes megjegyezni, hogy a mesterséges intelligenciában ma alkalmazott módszerek matematikailag nagyon hasonlóak a neokortexben működő mechanizmusokhoz.

Az "erős mesterséges intelligencia" (emberi szintű és azon túli mesterséges intelligencia) megvalósíthatóságával kapcsolatban gyakran felhozott másik ellenvetés, hogy az emberi agy nagymértékben használja az analóg számítástechnikát, míg a digitális módszerek természetüknél fogva nem képesek megismételni az analóg reprezentációk által megtestesített értékek fokozatosságát. Igaz, hogy egy bit vagy be van kapcsolva, vagy ki van kapcsolva, de a több bites szavak könnyedén több fokozatot képviselnek, és ezt bármilyen kívánt pontossággal megtehetik. Ez természetesen a digitális számítógépekben folyamatosan megtörténik. Az agyban az analóg információ (például a szinaptikus erősség) pontossága csak körülbelül egy szint a 256 szintből, amelyet nyolc bit ábrázolhat.

A [9. fejezetben](#) Roger Penrose és Stuart Hameroff ellenvetését idéztem, amely a mikrotubulusokról és a kvantumszámítástechnikáról szól. Emlékezzünk vissza, hogy állításuk szerint a neuronokban lévő mikrotubulus-szerkezetek kvantumszámítást végeznek, és mivel ezt a számítógépekben nem lehet elérni, az emberi agy alapvetően más és feltehetően jobb. Amint azt korábban állítottam, nincs bizonyíték arra, hogy a neuronális mikrotubulusok kvantumszámítást végeznek. Az emberek valójában nagyon rosszul oldják meg azokat a problémákat, amelyekben egy kvantumszámítógép kiválóan teljesítene (például a nagy számok faktorálását). És ha mindez igaznak bizonyulna, semmi sem zárna ki, hogy a kvantumszámítást a mi számítógépeinkben is alkalmazzuk.

John Searle híres az általa "kínai szobának" nevezett gondolat kísérlet bevezetéséről, amelyet részletesen tárgyalok a *The Singularity Is Near (A szingularitás közel van)* című könyvemben.⁹ Röviden: egy emberről van szó, aki kínai nyelven írásos kérdéseket fogad be, majd válaszol rájuk. Ehhez egy bonyolult szabálykönyvet használ. Searle azt állítja, hogy a férfinak nincs valódi

a kínai nyelv megértése, és nem "tudatosul" benne a nyelv (mivel nem érti sem a kérdéseket, sem a válaszokat), annak ellenére, hogy látszólag képes kínaiul válaszolni a kérdésekre. Searle ezt egy számítógéphez hasonlítja, és arra a következtetésre jut, hogy egy olyan számítógép, amely képes lenne kínaiul válaszolni a kérdésekre (lényegében átmenne egy kínai Turing-teszten), a kínai szobában lévő emberhez hasonlóan nem értene igazán a

nyelvet, és nem lenne tudatában annak, hogy mit csinál.

Searle érvelésében van néhány filozófiai bűvészmutatvány. Először is, az ember ebben a gondolat kísérletben csak a számítógép központi feldolgozó egységéhez (CPU) hasonlítható. Mondhatnánk, hogy a CPU nem érti igazán, hogy mit csinál, de a CPU csak egy része a szerkezetnek. Searle kínai szobájában az ember a szabálykönyvével *együtt* alkotja az egész rendszert. Ez a rendszer igenis érti a kínai nyelvet, különben nem lenne képes meggyőzően válaszolni a kínaiul feltett kérdésekre, ami sértené Searle feltételezését a gondolatkísérlethez.

Searle érvelésének vonzereje abból a tényből fakad, hogy ma már nehéz egy számítógépes programban valódi megértésre és tudatosságra következtetni. Az érvelésével azonban az a probléma, hogy saját érvelését magára az emberi agyra is alkalmazhatjuk. Minden neokortikális mintafelismerő - sőt, minden neuron és minden neuronális komponens - egy algoritmust követ. (Végül is ezek molekuláris mechanizmusok, amelyek természeti törvényeket követnek.) Ha arra a következtetésre jutunk, hogy az algoritmus követése nem egyeztethető össze az igazi megértéssel és a tudattal, akkor arra is következtetnünk kell, hogy az emberi agy sem rendelkezik ezekkel a tulajdonságokkal. Foghatjuk John Searle kínai szoba érvét, és egyszerűen csak kicserélhetjük a "szimbólumok manipulálása" szavakat "az interneuronális kapcsolatok és a szinaptikus erősségek manipulálásával", és máris meggyőző érvet kapunk arra, hogy az emberi agy nem érthet meg igazán semmit. Egy másik érvelési vonal a természet természetéből származik, amely sok megfigyelő számára új szent földdé vált. Michael Denton új-zélandi biológus (szül. 1943) például mély különbséget lát a gépek és a biológia tervezési elvei között. Denton azt írja, hogy a természeti entitások "önszerveződő önreferenciálisak, önreprodukálóak, kölcsönösek, önformálóak és holisztikusak." ¹⁰ Azt állítja, hogy ilyen biológiai formák csak biológiai folyamatok révén jöhetnek létre, és hogy ezek a formák ezáltal a létezés "megváltoztathatatlan, áthatolhatatlan és alapvető" valóságai, és ezért alapvetően más filozófiai kategóriát alkotnak, mint a gépek.

A valóságban, mint láttuk, a gépeket ugyanezen elvek alapján lehet megtervezni. A természet legintelligensebb lénye - az emberi agy - sajátos tervezési paradigmáinak megismerése pontosan az agy visszafejlesztési projektjének célja. Az sem igaz, hogy a biológiai rendszerek teljesen "holisztikusak", ahogy Denton fogalmaz, és fordítva, a gépeknek sem kell teljesen modulárisnak lenniük. A természetes rendszerekben, különösen az agyban, egyértelműen azonosítottuk a funkcionalitás egységeinek hierarchiáját, és a mesterséges intelligencia rendszerek hasonló módszereket alkalmaznak.

Úgy tűnik számomra, hogy sok kritikus nem lesz elégedett addig, amíg a számítógépek nem mennek át rutinszerűen a Turing-teszten, de még ez a küszöb sem lesz egyértelmű. Kétségtelen, hogy lesznek viták arról, hogy az állítólagos Turing-tesztek, amelyeket már elvégeztek, érvényesek-e. Valószínűleg én is azok közé a kritikusok közé fogok tartozni, akik a korai állításokat ilyen irányban ócsárolják. Mire a Turing-tesztet teljesítő számítógépek érvényességéről szóló viták lecsillapodnak, a számítógépek már régen túlszárnyalják az emberi intelligenciát.

A hangsúlyt a "nem feljavított" szóra helyezem, mert a feljavítás pontosan az oka annak, hogy létrehozzuk ezeket az "elme gyermekeket", ahogy Hans Moravec nevezi őket. ¹¹

Az emberi szintű mintafelismerés kombinálása a számítógépek eredendő sebességével és pontosságával nagyon erős képességeket fog eredményezni. De ez nem a Marsról érkező intelligens gépek idegen inváziója - mi azért hozzuk létre ezeket az eszközöket, hogy okosabbá tegyük magunkat. Azt hiszem, a legtöbb megfigyelő egyet fog érteni velem abban, hogy ez az, ami az emberi fajban egyedülálló: Azért építjük ezeket az eszközöket, hogy kiterjesszük a saját hatókörünket.

A kép elég sivár, uraim... A világ éghajlata változik, az emlősök átveszik az uralmat, és mindannyiunk agya diónyi méretű.

-Dinoszauruszok beszélnek, Gary Larson: *A túlsó oldal* című könyvében.

Az intelligencia úgy határozható meg, mint a korlátozott erőforrásokkal történő problémamegoldás képessége, amelyben az egyik legfontosabb ilyen erőforrás az idő. Így az olyan problémák gyorsabb megoldásának képessége, mint például az élelemkeresés vagy a ragadozók elkerülése, az értelem nagyobb erejét tükrözi. Az intelligencia azért fejlődött ki, mert hasznos volt a túlélés szempontjából - ez a tény nyilvánvalónak tűnhet, de nem mindenki ért egyet vele. Ahogyan a mi fajunk gyakorolja, nemcsak azt tette lehetővé, hogy uraljuk a bolygót, hanem azt is, hogy folyamatosan javítsuk az életminőségünket. Ez utóbbi szintén nem mindenki számára nyilvánvaló, hiszen manapság széles körben elterjedt az a vélekedés, hogy az élet csak rosszabbodik. Egy májusban közzétett Gallup-felmérésből például kiderült⁴, 2011., hogy csak "a 44 százaléknyi

Az amerikaiak hittek abban, hogy a mai fiataloknak jobb életük lesz, mint a szüleiknek."¹

Ha az általános tendenciákat nézzük, nem csak az emberi életet a várható élettartam az elmúlt évezredben megnégyszereződött (és az elmúlt két évszázadban több mint megduplázódott),² de az egy főre jutó GDP (állandó folyó dollárban kifejezve) az 1800-as száz dollárról napjainkra több ezer dollárra emelkedett, és a fejlett világban még ennél is markánsabb tendenciákat mutat.³ Egy évszázaddal ezelőtt csak néhány demokrácia létezett, míg ma ezek a demokráciák a normák. Ha valaki történelmi távlatból szeretné látni, hogy mennyit fejlődtünk, javasolom, hogy olvassa el Thomas Hobbes *Leviatánját* (1651), amelyben az "ember életét" úgy írja le, mint "magányos, szegény, csúnya, brutális és rövid". Modern szemléletet nyújt Peter Diamandis, az X-Prize Alapítvány alapítója (és velem együtt a Singularity University társalapítója) és Steven Kotler tudományos író *Abundance* (2012) című könyve, amely dokumentálja, hogy napjainkban az élet minden dimenzióban folyamatosan javul. Steven Pinker nemrégiben megjelent *The Better Angels of Our Nature (Természetünk jobb angyalai): Why Violence Has Declined* (2011) című könyvében aprólékosan dokumentálja az emberek és népek közötti békés kapcsolatok folyamatos növekedését. Martine Rothblatt (1954-ben született) amerikai ügyvéd, vállalkozó és író a polgári jogok folyamatos javulását dokumentálja, és megállapítja például, hogy néhány évtized alatt az azonos neműek házassága a világon sehol sem volt jogilag elismert, most pedig egyre több országban elfogadott.⁴

Az egyik fő oka annak, hogy az emberek úgy gondolják, hogy az élet egyre rosszabb, az, hogy a világ problémáival kapcsolatos információink folyamatosan javulnak. Ha ma valahol a bolygón harc folyik, azt szinte úgy éljük meg, mintha mi is ott lennénk. A második világháború idején emberek tízezrei pusztulhattak el egy csatában, és ha a közönség

egyáltalán láthatta, akkor azt hetekkel később egy szemcsés filmhíradóban láthatta a moziban. Az első világháború alatt egy kis elit olvashatott a konfliktus előrehaladásáról az újságban (képek nélkül). A tizenkilencedik században szinte senki sem tudott időben hozzáférni a hírekhez.

Az intelligenciánknak köszönhető faji fejlődésünk tükröződik tudásunk fejlődésében, amely magában foglalja a technológiánkat és a kultúránkat is. Különböző technológiáink egyre inkább információs technológiákká válnak, amelyek természetüknél fogva exponenciális módon fejlődnek tovább. Az ilyen technológiák révén vagyunk képesek az emberiség nagy kihívásainak megoldására, mint például az egészséges környezet fenntartása, a növekvő népesség számára az erőforrások biztosítása (beleértve az energiát, az élelmet és a vizet), a betegségek legyőzése, az emberi élettartam jelentős meghosszabbítása és a szegénység felszámolása. Csak az intelligens technológiával való bővítésünkkel tudjuk kezelni az e kihívások kezeléséhez szükséges komplexitás nagyságrendjét.

Ezek a technológiák nem egy intelligens invázió előhadát jelentik, amely versenyezni fog velünk, és végül kiszorít minket. Amióta csak felkaptunk egy botot, hogy elérjünk egy magasabb ágat, azóta használjuk eszközeinket, hogy kiterjesszük hatókörünket, mind fizikailag, mind szellemileg. Az, hogy ma elővehetünk egy eszközt a zsebünkben, és néhány billentyűleütéssel hozzáférhetünk az emberi tudás nagy részéhez, túlmutat mindenben, amit a legtöbb megfigyelő csak néhány évtizeddel ezelőtt elképzelhetett. A zsebemben lévő "mobiltelefon" (a kifejezés azért van idézőjelben, mert ez sokkal több, mint egy telefon) milliószor olcsóbb, mégis ezerszer erősebb, mint az MIT összes diákjának és professzorának közös számítógépe, amikor én még egyetemista voltam. Ez az elmúlt negyven év alatt több milliárdszoros ár/teljesítmény növekedést jelent, és ezt a növekedést a következő huszonöt évben ismét látni fogjuk, amikor az, ami korábban egy épületben, most pedig a zsebünkben elfér, egy véresejtben fog elférni.

Így fogunk összeolvadni az általunk létrehozott intelligens technológiával. A véráramunkban lévő intelligens nanorobotok fogják biológiai testünket sejtszinten és molekuláris szinten egészségesen tartani. A kapillárisokon keresztül noninvazív módon jutnak majd be az agyunkba, és kölcsönhatásba lépnek biológiai neuronjainkkal, közvetlenül kiterjesztve intelligenciánkat. Ez nem olyan futurisztikus, mint amilyennek hangzik. Már léteznek véresejt méretű eszközök, amelyek képesek az állatokban az I. típusú cukorbetegséget gyógyítani, vagy a véráramban lévő rákos sejteket felismerni és elpusztítani. A gyorsuló megterülés törvénye alapján ezek a technológiák három évtizeden belül egymilliárdszor erősebbek lesznek, mint ma.

Az általam használt eszközöket és a számítástechnikai erőforrások felhőjét, amelyekhez virtuálisan kapcsolódnak, már most is önmagam kiterjesztésének tekintem, és nem érzem magam teljesnek, ha elszakadok ezektől az agyi kiterjesztőktől. Ezért volt olyan figyelemre méltó a Google, a Wikipedia és több ezer más weboldal egynapos sztrájkja a SOPA (Stop Online Piracy Act) ellen 2012. január 18-án: Úgy éreztem, mintha az agyam egy része sztrájkolna (bár én és mások is megtaláltuk a módját, hogy hozzáférjünk ezekhez az online forrásokhoz). Ez egyben e webhelyek politikai erejének lenyűgöző demonstrációja is volt, mivel a törvényjavaslatot - amely úgy tűnt, hogy ratifikáció felé tart - azonnal megölték. De ennél is fontosabb, hogy megmutatta, milyen alaposan kiszerveztük már gondolkodásunk egy részét a számítástechnikai felhőbe. Ez már része annak, hogy kik

vagyunk. Ha egyszer már rutinszerűen intelligens, nem biológiai intelligenciával rendelkezünk az agyunkban, akkor ez a bővítés - és a felhő, amelyhez kapcsolódik - exponenciálisan tovább fog növekedni.

Az intelligencia, amelyet az agy visszafejtéséből fogunk létrehozni, hozzáfér a saját forráskódjához, és képes lesz önmagát gyorsan fejleszteni egy gyorsuló iteratív tervezési ciklusban. Bár a biológiai emberi agyban, mint láttuk, jelentős plaszticitás van, viszonylag kötött felépítéssel rendelkezik, amelyet nem lehet jelentősen módosítani, valamint korlátozott kapacitással. Nem tudjuk 300 millió mintafelismerőjét mondjuk 400 millióra növelni, hacsak nem nem biológiai úton tesszük ezt. Amint ezt elérjük, nem lesz okunk megállni egy adott képességi szinten. Eljuthatunk egymilliárd vagy akár trillió mintafelismerőig is. A mennyiségi fejlődésből minőségi előrelépés következik. A *Homo sapiens* legfontosabb evolúciós fejlődése mennyiségi volt: a nagyobb homlok kialakulása, hogy több neokortexet tudjon befogadni. A nagyobb neokortikális kapacitás lehetővé tette, hogy ez az új faj magasabb fogalmi szinteken alkosson és gondolkodjon, ami a művészet és a tudomány valamennyi változatos területének létrejöttét eredményezte. Ahogy egyre több neokortexet adunk hozzá nem biológiai formában, egyre magasabb minőségi absztrakciós szintekre számíthatunk.

Irvin J. Good brit matematikus, Alan Turing kollégája 1965-ben azt írta, hogy "az első ultraintelligens gép az utolsó találmány, amelyet az embernek valaha is meg kell alkotnia". Egy ilyen gépet úgy definiált, mint amely képes felülmúlni "bármely ember bármilyen okos szellemi tevékenységét", és arra a következtetésre jutott, hogy "mivel a gépek tervezése az egyik ilyen szellemi tevékenység, egy ultraintelligens gép még jobb gépeket tudna tervezni; ekkor kétségtelenül "intelligencia-robbanás" következne be".

Az utolsó találmány, amelyet a biológiai evolúciónak létre kellett hoznia - a neokortex - elkerülhetetlenül elvezet az utolsó találmányhoz, amelyet az emberiségnek létre kell hoznia - a valóban intelligens gépekhez -, és az egyik tervezése inspirálja a másikat. A biológiai evolúció folytatódik, de a technológiai evolúció milliószor gyorsabban halad, mint az előbbi. A gyorsuló megtérülés törvénye szerint az évszázad végére képesek leszünk arra, hogy a fizika törvényei alapján a számítást a lehetőségek határán hozzuk létre, ahogyan azt a fizika a következő területeken alkalmazza

számítás.⁵ Az így szervezett anyagot és energiát "komputroniumnak" nevezzük, amely fontról fontra sokkal erősebb, mint az emberi agy. Ez nem csak nyers számítás lesz, hanem intelligens algoritmusokkal lesz átítva, amelyek az ember-gép tudás egészét alkotják. Idővel a galaxisunk aprócska szegletében lévő, erre a célra alkalmas tömeg és energia nagy részét komputroniummá alakítjuk át. Aztán, hogy a gyorsuló megtérülés törvényét fenntartsuk, ki kell majd terjeszkednünk a galaxis és az univerzum többi részére.

Ha a fénysebesség valóban megállíthatatlan határ marad, akkor a világegyetem benépesítése hosszú időbe telik, tekintve, hogy a Földhöz legközelebbi csillagrendszer négy fényévre van. Ha léteznek akár csak finom eszközök is e korlát megkerülésére, intelligenciánk és technológiánk elég erős lesz ahhoz, hogy kihasználjuk őket. Ez az egyik oka annak, hogy a közelmúltban felröppent hír, miszerint a svájci-francia határon lévő CERN gyorsítótól a közép-olaszországi Gran Sasso laboratóriumig 730 kilométert megtett müonok a fénysebességnél gyorsabban mozogtak, potenciálisan olyan fontos hír volt. Úgy

tűnik, hogy ez a konkrét megfigyelés téves riasztásnak bizonyult, de vannak más lehetőségek is arra, hogy ezt a határt megkerüljük. Még csak nem is kell túllépnünk a fénysebességet, ha az általunk ismert három dimenzió túl térbeli dimenziókon keresztül rövidebb utakat találunk más, látszólag távoli helyekre. Az, hogy képesek leszünk-e túllépni vagy más módon megkerülni a fénysebességet mint korlátot, a huszonkettedik század elején az ember-gép civilizáció stratégiai kulcskérdése lesz.

A kozmológusok arról vitatkoznak, hogy a világ tűzben (az ősrobbanáshoz hasonló nagy roppanásban) vagy jégben (a csillagok halála, ahogyan azok örök tágulásba terjeszkednek) ér-e véget, de ez nem veszi figyelembe az intelligencia erejét, mintha annak megjelenése csak egy szórakoztató mellékszál lenne a világegyetemet jelenleg uraló nagy égi mechanikához képest. Mennyi időbe telik, amíg intelligenciánkat nem biológiai formában elterjesztjük az egész világegyetemben? Ha képesek vagyunk túllépni a fénysebességet - ami kétségtelenül egy nagy "ha" - például a téren átvezető féreglyukak segítségével (amelyek összhangban vannak a fizika jelenlegi megértésével), akkor ez néhány évszázadon belül megvalósulhat. Máskülönben valószínűleg sokkal tovább tartana. Mindkét forgatókönyv szerint az univerzum felébresztése, majd a sorsáról való intelligens döntés az emberi intelligencia nem biológiai formában történő beoltásával a mi sorsunk.