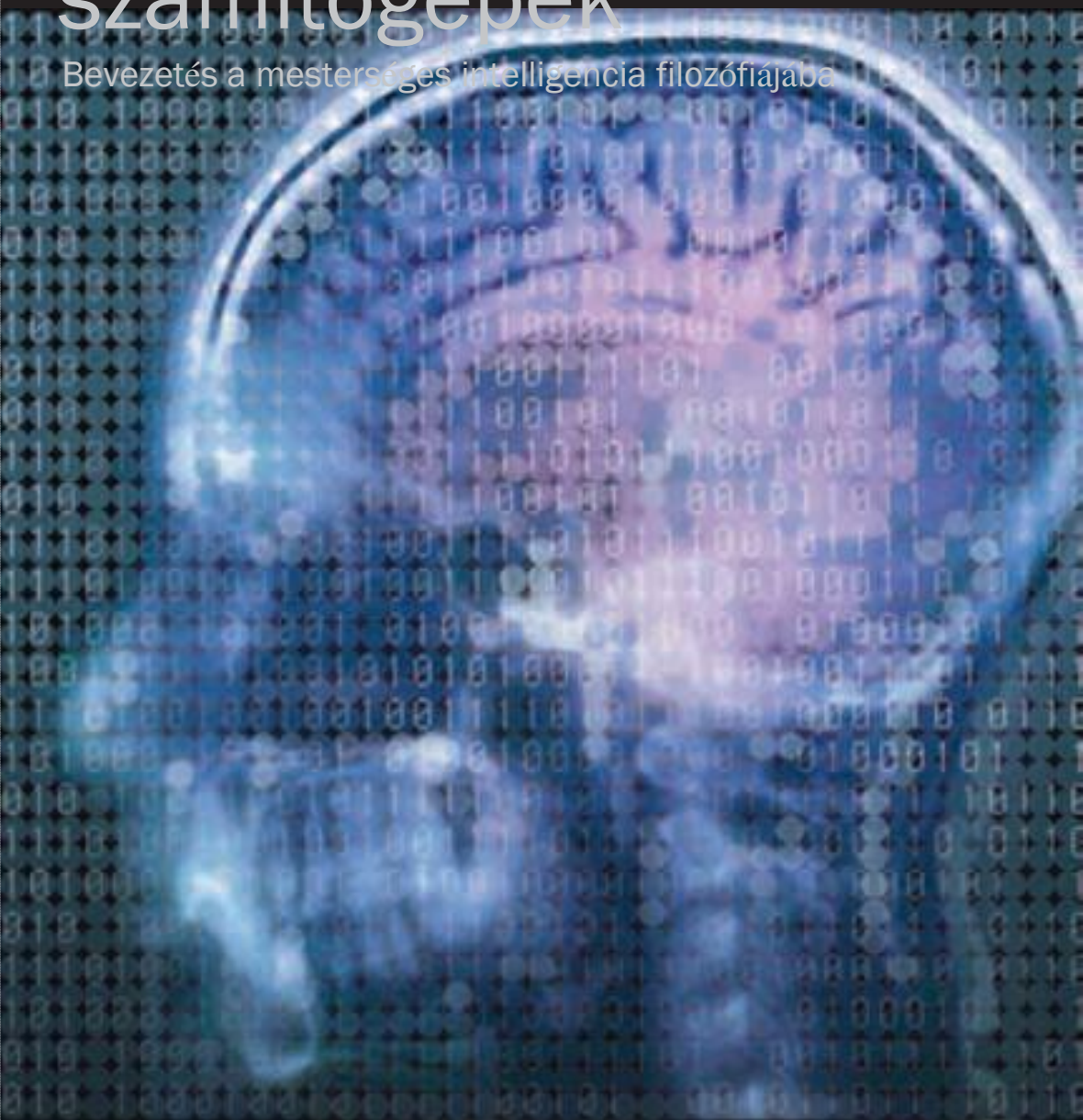


Elmék és számítógépek

Bevezetés a mesterséges intelligencia filozófiájába



Matt Carter

ELMÉK ÉS SZÁMÍTÓGÉPEK

G-nek,
aki segített

és

Sue-

nak
akik nélkül . . .

ÉSZAKOK
ÉS
SZÁMÍTÓSZÁMOK

BEVEZETÉS A MESTERSÉGES
INTELLIGENCIAFILOZÓFIÁJÁBA

Matt Carter

EDINBURGH UNIVERSITY PRESS

© Matt Carter, 2007

Edinburgh University Press Ltd
22 George Square, Edinburgh

Times betűkészlet
Servis Filmsetting Ltd, Manchester, és
nyomtatta és kötötte Nagy-Britanniában a Cromwell Press,
Trowbridge, Wilts

A könyv CIP-nyilvántartása a British
Library-nél érhető el.

ISBN 978 978 0 7486 2098 2 (kemény kötés)
ISBN 978 978 0 7486 2099 9 (kartonozott)

Matt Carter jobb oldala
hogy e mű szerzőjeként azonosítható
legyen, a következőkkel összhangban
érvényesítették
a szerzői jogról, formatervezési mintákról és szabadalmakról szóló
1988. évi törvény.

TARTALOM

	önetyilvánítás	Kösz
1	Bevezetés	1
2	Dualizmus	4
2.1	Anyag dualizmus	4
2.2	Kartéziánus dualizmus	5
2.3	Pozitív érvek a karteziánus dualizmus mellett	5
2.3.1	A érv	vallási
2.3.2	Az introspektív megjelenés érve	7
2.3.3	Az alapvető tulajdonságokravonatkozó	érv
2.4	Érvek a karteziánus dualizmus ellen	9
2.4.1	A más problémája	elmék
2.4.2	Ockham borotvája	10
2.4.3	A kölcsönhatás problémája	11
2.5	Egyéb dualizmusok	12
2.5.1	Párhuzamosság	12
2.5.2	Alkalmasság	13
2.5.3	Epifenomenalizmus	13
2.6	Anomális monizmus	14
3	Behaviourizmus	15
3.1	Korai empirikus pszichológia	15

- 3.2 Fiziológiai pszichológia16
 - 3.3 Introspekcionista pszichológia17
 - 3.4 Pszichológiai viselkedélmélet20
 - 3.5 Filozófiai viselkedélmélet23
 - 3.6 Ellenvetések a filozófiai viselkedélmélettel szemben24
- 4 Neuroanatómia27
 - 4.1 Makro-neuroanatómia27
 - 4.2 Mikro-neuroanatómia32

5	Ausztrál materializmus	35
	5.1A tudat oksági elmélete	36
	5.2Azidentitáselmélet	37
	5.3Aausztrál materializmus elleni érvek	38
	5.4AmitMary nem tudott	42
6	Funkcionalizmus	44
	6.1 Funkcionálismeghatározás	44
	6.2AFekete doboz elmélet	45
	6.3Qualiakifogások	48
7	Formális rendszerek	52
	7.1Effektivitás	53
	7.2Államokés szabályok	57
	7.3Specifikáció	58
	7.4Generáció és levezetés	61
	7.5Generációsfák	64
	7.6Formalitásés izomorfizmus	67
8	Számíthatóság	70
	8.1Regisztráltgépek	70
	8.2Programok	71
	8.3Program futtatása	73
	8.4Kiszámítás	75
	8.5Kiszámíthatófüggvények	76
	8.6Építésiprogramok	79
9	Univerzális gépek	85
	9.1Church/Turingtézis	86
	9.2Gödelkódolás	88
	9.3AUniverzális gép	92
10	Computationalism	94
	10.1 Ami a komputationalizmus nem	95
	10.2 Szoftver és Wetware	99
	10.3 Variáció	101
	10.4 Tanulás	103
	10.5 Kreativitás	106
	10.6 A mentalitás tulajdonítása	108
	11Ker	113
esés		
	11.1 Felülről lefelé, alulról felfelé	113
	11.2 Széleskörűség kontra mélység	115

11.3	Heurisztikus keresés	117
		vii
12Ga		122
mes		
12.1	Egy egyszerű játék	122
12.2	Minimax	125
12.3	Metszés	127
12.4	Emberek kontra számítógépek	128
13	Gépi érvelés	132
	13.1 Logika és dedukció	132
	13.2 Feltételeesség és predikció	134
	13.3 Rokonság	137
	13.4 Expert Systems	141
14	Gépek és nyelv	145
	14.1 Nyelvi tolmácsolás	145
	14.2 Generatív nyelv tan	149
	14.3 Kifejezésszerkezeti fák	150
	14.4 Számítógépes nyelv	152
15	Emberi érvelés	154
15.1	Logikusan követni	156
15.2	Racionális teljesítmény	157
15.3	Mental modellek	160
15.4	Magyarázó teher	161
16	Emberi nyelv	164
	16.1 Előrejtő fonémák	165
	16.2 Sonoráns fonémák	167
	16.3 Allofonok és fonetikai megvalósítás	170
	16.4 Az első nyelv elsajátítása	172
	16.5 Nyelvés szabályok	173
17	Jelentése	175
	17.1 Akínai szoba	175
	17.2 Szintaktika és szemantika	177
18	Képviselés	181
	18.1 Szándékosság	181
	18.2 Kategóriák és tartalom	183
	18.3 Szimbólumok és minták	184
	18.4 Kognitív architektúra	185
19	Mesterséges neurális hálózatok	187
	19.1 Kapcsolati architektúra	187

viii

19.3	Beszéd szintetizálása	191
19.4	Tanulás	196
19.5	Mintafelismerés	198
19.6	Két paradigma?	199
19.7	Ez csak egy modell	200
20	Elmék és	számítógépek202
20.1	Tudatosság	202
20.2	Személyes	identitás203
20.3	Érzelmek	205
20.4	Számítógépek	elmével206
I. függelék:	Javaslatok további	olvasmányokhoz207
II. függelék:		Fogalomtár211
		Index218

KÖSZÖNETNYILVÁNÍTÁS

Szeretném kifejezni hálámat mindazoknak, akik közvetlenül és közvetve részt vettek e könyv elkészítésében.

Köszönöm a Queenslandi Egyetem kognitív tudományok programjának oktatóinak, akik a huszadik század utolsó éveiben a Queenslandi Egyetem kognitív tudományok programjában dolgoztak, hogy elkötelezettséget ébresztettek bennem a tudományágak közötti elemzés fontossága iránt. Hálám kiterjed továbbá minden tanítómra - mind a filozófián belül, mind azon kívül.

Köszönöm az éveken át tartó egyetemi hallgatóknak is, akik figyelemmel kísérték az oktatásomat. Különösen hálás vagyok a 2002-es "Elmék és gépek" osztályomnak, amiért először is inspirálták ezt a tankönyvet, és a 2006-os osztályomnak, amiért elolvasták és kommentálták az itt található anyagot.

Köszönöm a Melbourne-i Egyetem Filozófia Tanszékének, ahol e könyv megírása alatt dolgoztam, és a kiváló office staffnak.

Köszönöm mindenkinek az EUP-nál, hogy kiadták ezt a kötetet, és hogy ilyen kellemes volt velük dolgozni. Külön köszönet Jackie Jonesnak a projekt iránti kezdeti lelkesedéséért.

Köszönöm minden barátomnak a támogatást és a megértést, különösen FB-nek, Wayne-nek és Eloise-nak, hogy elviselték a számtalan vacsorai látogatást, valamint Lesternek és Christie-nek a barátságon felüli segítségüket.

Köszönöm szépen Graham Priestnek, aki nélkül ez a könyv nem született volna meg.

Köszönjük Mia és Linus imádnivaló volt, és millió köszönet Sue-nak, hogy csodálatos volt.

1. FEJEZET

BEVEZETÉS

Ez a könyv az elméről szól. A számítógépekről is szól. A könyv középpontjában az elme és a számítógépek közötti kapcsolat vizsgálata áll.

Az elképzelés, hogy egy nap talán képesek leszünk olyan tárgyakat létrehozni, amelyeknek ugyanolyan értelemben van elméjük, mint nekünk van elménk, nem új keletű. Mary Shelley Frankenstein szörnyének megalkotása óta szerepel szórakoztató és ijesztő regényekben.

A huszadik század elejének és közepének klasszikus sci-fijében ezt az elképzelést általában a "mechanikus emberek" vagy *robotok* fogalmával fejezték ki - a cseh *robata* szóból, amely nagyjából a feudális *corvée* kifejezéssel fordítható le, amely kifejezés a hűbérúrnak nyújtott fizetetlen munkára utal.

A modernebb fikciókban a mechanikus elme gondolata átadta helyét a számítógépes *mesterséges intelligencia* ma már általános fogalmának. A mesterséges intelligencia tényleges kifejlesztésének lehetősége azonban nem csupán a fejlett technológia kérdése. Ez alapvetően *filozófiai* kérdés.

Ez az a kérdés, amellyel e kötetben központi szerepet kap. Ahhoz, hogy megalapozottan mérlegelhessük a mesterséges intelligencia lehetőségét, számos kapcsolódó kérdésre kell választ adnunk.

Először is azt fogjuk megkérdezni, hogy mi is az emberi elme. A kétezredik században az elme filozófiai elméleteinek sora váltotta egymást, és a jelenleg uralkodó elméletben csúcsosodott ki, amely a mesterséges intelligencia lehetőségét is magában foglalja. Első célunk, amelynek az 1-10. fejezetet szenteljük, ennek az elméletnek a világos megfogalmazása.

Az elmélettel való filozófiai felelős foglalkozás megköveteli, hogy pontosan megértsük, mi is a számítógép. Következésképpen három fejezetet fogunk azzal tölteni, hogy kidolgozzuk a számítás szigorú technikai beszámolóját. Bár ez az anyag technikai jellegű, a

A bevezetés lassú és szelíd, és könnyen hozzáférhető a matematikai vagy informatikai háttérrel nem rendelkező olvasó számára is.

A célelméletünkhöz vezető úton áttekintjük a rendelkezésre álló filozófiai elméleteket, mérlegelve mindegyiknek az érdemeit és hibáit. Ezáltal átfogó bevezetést nyújtunk az elmefilozófiába.

Útközben néhány empirikus kitérőt is teszünk. Elmondjuk az empirikus pszichológia kialakulásának történetét, és egy fejezetet a funkcionális neuroanatómia kezdetleges megértésének kidolgozásával fogunk tölteni.

Ha már rendelkezünk a célul kitűzött elmélet filozófiai megértésével, a könyv hátralévő részében értékelni fogjuk azt. Látni fogjuk, hogy az empirikus diszciplínák anyagának széles skálája jelentősen befolyásolja az elmélet tarthatóságát. Mint ilyen, ez a könyv átfogóan a multidiszciplináris elemzés gyakorlata.

Két olyan mentális képességre fogunk összpontosítani, amelyek jellegzetesen emberi tulajdonságok - a gondolkodási képességünk és a nyelvi képességünk. A 11-20. fejezetben mindenekelőtt az lesz a célunk, hogy összevessük azt, amit az emberi racionális és nyelvi képességekről tudunk, az ezeket számítással megvalósító módszerekkel.

Látni fogjuk, hogyan programozhatók a számítógépek stratégiai játékra, az ismert információk alapján új helyzetekből való következtetésre, valamint a nyelvi termelésben és megértésben szerepet játszó bizonyos funkciókra. Ennek során megismerkedünk a nyelvészet némi bevezető anyagával és egy kis formális logikával, és érintjük a kognitív pszichológia némi anyagát is.

A könyv utolsó fejezeteiben a jelentés és a reprezentáció fogalmával kapcsolatos haladóbb filozófiai anyagot vizsgáljuk meg. Végül bemutatjuk a mesterséges neurális hálózatokat, és megnézzük, hogyan lehet őket alkalmazni a mesterséges intelligenciára való törekvésben - ismét különös tekintettel a racionális és nyelvi funkciókra.

Összességében a filozófia, a ~~pszich~~ nyelvészet, az idegtudomány és az informatika - a kognitív tudományt alkotó tudományágak - anyagát fogjuk megvizsgálni. Várható, hogy a legtöbb olvasó számára az anyag egy része könnyebben, más része pedig kevésbé lesz megközelíthető; mindazonáltal végig arra törekedtem, hogy a bevezető olvasó számára maximálisan hozzáférhető legyen.

A 11. fejezettől kezdve minden egyes fejezet egy olyan témával foglalkozik, amely minden joggal megérdemelne egy külön kötetet. Ezért a terjedelem nem mindenre kiterjedő, és a hozzáférhetőség érdekében gyakran egyszerűsítettem a magyarázatokat. Vannak javaslatok a további

a könyv végén található olvasmányt azoknak az olvasóknak, akik tovább akarják mélyíteni az általunk tárgyalt témák megértését.

A vonatkozó kérdések átfogó lefedése azonban itt nem elsődleges szempont. Fő célunk az elme filozófiai elméletének kidolgozása és értékelése, amely lehetővé teszi a mesterséges intelligencia lehetőségét. A könyv végére az olvasónak olyan megalapozott helyzetben kell találnia magát, amelyből kiindulva megalapozott döntéseket hozhat a mesterséges intelligencia kifejlesztésének lehetőségével kapcsolatban. Ez a könyv emellett szilárd alapot biztosít a kognitív tudományokkal való filozófiai felelős foglalkozáshoz is.

Az elme rendelkezésre álló filo- szofikus elméletek terében tett körutazásunkat most egy olyan elmélettel kezdjük, amelyet a legtöbb ember implicit módon és preteoretikusan elfogad: a dualizmussal.

2. FEJEZET

DUALIZMUS

Az elme rendelkezésre álló elméleti elméleteinek vizsgálatát a karteziánus dualizmussal kezdjük. Legalább két jó okunk van rá, hogy ezt tegyük. Az egyik az, hogy az elmeelméletek bemutatása a következő fejezetekben nagyjából kronologikus lesz, és - legalábbis ami a modern filozófiát illeti - a karteziánus dualizmussal kezdeni annyit jelent, mint az elején kezdeni.

A másik ok az, hogy az emberek elmével és testtel kapcsolatos preteoretikus intuíciói általában véve dualista, különösen pedig karteziánus dualista nézeteket vallanak. Ha elolvasta ezt a fejezetet, kérdezze meg barátait és családtagjait az intuícióikról, és erős a gyanúm, hogy ugyanezt fogja tapasztalni.

Sajnos, bár jó kiindulópont, a karteziánus dualizmus filozófiai nehézségekkel küzd. Kezdjük hát hozzá ahhoz a feladathoz, hogy pontosan tisztázzuk, mi mellett kötelezi el magát a karteziánus dualista, és mik a problémák az elmélettel.

2.1 SZUBSZTANCIA DUALIZMUS

A szubsztancia dualizmus egy metafizikai nézet. Ez az a nézet, amely szerint az egység két különböző típusú anyagból - két metafizikailag elkülönülő szubsztanciából - áll. Mint ilyen, az anyagdualizmus egy bizonyos *ontológia* melletti elkötelezettség - egy olyan ontológia, amely a világegyetemet anyagi és nem anyagi szubsztanciákból állónak tekinti. A dualisták szerint a fizikai világot alkotó anyagi anyag mellett a nem fizikai, immateriális anyagokat is figyelembe kell venni.

Óvatosnak kell lennünk, amikor különbséget teszünk a matematikai és az immateriális között. Például az elektromágneses sugárzás, bár bizonyos értelemben anyagtalan, mégis anyagi. A ~~fi~~ világ része - olyasvalami, amiről a fizikától elvárnánk, hogy számot adjon nekünk. Az anyagi és az immateriális közötti különbségtétel nem csupán egyszerű különbségtétel olyan dolgok között, amelyekbe beleütközhetünk a világban.

sötétben (székek és asztalok) és olyan dolgok, amiket nem (hő, fény és hang). Inkább a fizika hatáskörébe tartozó dolgok (székek, asztalok, hő, fény, hang) és a dualista szerint a fizika hatókörén kívül eső dolgok közötti különbségtételről van szó.

Ezt elvontan nehéz felfogni, ezért vizsgáljuk meg az anyagdualizmus egy sajátos fajtáját, és nézzük meg, miért lehetünk hajlamosak azt állítani, hogy a világegyetemben léteznek nem fizikai, anyagtalan anyagból álló tárgyak.

2.2 KARTÉZIÁNUS DUALIZMUS

A karteziánus dualizmus az elméről, a testről és a köztük lévő kapcsolatról alkotott nézet. Ez a dualizmus egy sajátos fajtája, amely nevét eredeti képviselőjéről, René Descartes-ról kapta. Lényegében azt a nézetet képviseli, hogy míg a test anyagi tárgy, addig az elme nem az. A karteziánus dualista szerint az elme teljes egészében anyagtalan stuffból áll. Mint ilyen, a karteziánus dualizmus egyértelműen egyfajta szubsztancialista dualizmus, mivel a karteziánus dualista olyan ontológia mellett kötelezte el magát, amely mind anyagi, mind immateriális szubsztanciákat megenged.

A karteziánus dualizmus megkülönböztető jegye az elme-test dualizmus más fajtái között az, hogy a karteziánus dualista szerint az elme és a test kauzális viszonyba lép egymással: az elme okozza, hogy a testben dolgok történnek, és a test okozza, hogy az elmében dolgok történnek. Más szóval, az anyagtalan elme és az anyagi test *kölcsönhatásban áll egymással*. A karteziánus dualizmust emiatt interakcionista dualizmusnak is nevezik.

A karteziánus dualista a következő négy tétel mellett kötelezi el magát:

- [D1] A test teljes egészében anyagi anyagból áll
- [D2] Az elme teljes egészében anyagtalan anyagból áll
- [D3] A testnek kauzális hatása van az elmére.
- [D4] Az elme kauzális hatással van a testre.

Amint azt rövidesen látni fogjuk, nagyon diffikult mind a négy tételt fenntartani. Mielőtt azonban a karteziánus dualizmussal szembeni ellenvetéseket vizsgálnánk meg, először nézzük meg a mellette szóló érveket.

2.3 POZITÍV ÉRVEK A KARTÉZIÁNUS DUALIZMUS MELLETT

Számos oka van annak, hogy valaki a karteziánus dualizmust támogatja. A három legerősebb érvet fogom megvizsgálni mellette: a következők.

a vallási érv, az introspektív megjelenésből származó érv és a lényegi tulajdonságokból származó érv.

2.3.1 A VALLÁSIÉRV

Ez talán a leggyakrabban hangoztatott érv a karteziánus dualizmus mellett.

Számos vallás - köztük a kereszténység - feltételez egy túlvilági életet, és jutalmat ígér a túlvilágon egy bizonyos normatív kódex szerint élőknek. Ezzel szemben büntetéssel fenyegetnek a túlvilágon, ha valaki nem ezen előírások szerint él. De kérdezzük meg magunktól: ki az, akit megjutalmaznak vagy megbüntetnek?

Az ilyen vallások az örökkévaló, megváltoztathatatlan, anyagtalan lélekről beszélnek, amelyről azt állítják, hogy fontos értelemben az egyén alkotóeleme. Ez az örök lélek az, amely a túlvilágon élvezheti a jutalmakat vagy elszenvedheti a büntetéseket.

Ahhoz, hogy a jutalmazás és a büntetés fogalma alkalmazható legyen - és ahhoz, hogy a vonatkozó hiedelmek az egyéneket a megfelelő cselekvésre ösztönözzék -, annak kell lennie, hogy a jutalmazott vagy büntetett dolog ugyanaz a dolog, ami az erkölcsi cselekvésért felelős.

Más szóval, a jutalmazott vagy büntetett dolognak egyszerűen ugyanannak a dolognak *kell* lennie, amely a világban jár, és bizonyos döntéseket hoz, és bizonyos módon cselekszik. A léleknek egyszerűen egyenértékűnek kell lennie az elmével. Végül is mi értelme van annak, hogy egy entitást egy másik entitás tetteiért vagy vétkeiért jutalmazunk vagy büntessünk? És miért kellene egyáltalán foglalkoznom azzal, hogy egy bizonyos kódex szerint cselekedjek, ha egy nagyon fontos értelemben nem *én* vagyok az, aki élvezni fogja az ígért jutalmat, vagy elszenvedi a fenyegető büntetést?

A karteziánus dualisták az "elme" és a "lélek" szinonim fogalmak. Míg az életben az elme/lélek egy adott testhez viszonyítva áll. A túlvilágon az anyagtalan elme/lélek elhagyja a testet, hogy önálló létet kezdjen, és élvezze vagy élvezze az anyagi testet öltött élete során végzett cselekedeteinek jutalmát.

Tehát, amennyiben valaki előzetesen elkötelezett egy ilyen vallási tan mellett, akkor az anyagdualizmus mellett is elkötelezettnek kell lennie, mivel az ilyen tanok egy anyagtalan lelket követelnek meg. Továbbá, a karteziánus dualizmus mellett is el kell köteleznünk magunkat, mivel az elmének/léleknek *kell okoznia* a test cselekedeteit, és ezáltal felelősnek kell lennie azokért.

Ami az érveket illeti, ez sajnos nem túl jó érv. Nem ad semmilyen független indokot a karteziánus dualizmus támogatására. Azt mutatja, hogy a karteziánus dualizmus melletti elköteleződés

bizonyos vallási meggyőzések egyenes következménye. Ez egyszerűen azt jelenti, hogy ezek a vallási meggyőzések a karteziánus dualizmussal együtt állnak vagy bukhatnak. Ha valaki nem rendelkezik előzetesen ilyen vallási meggyőzésekkel, akkor a vallási érvnek teljesen hiányzik a meggyőző ereje.

2.3.2 AZ ÉRV AZ INTROSPEKTÍV MEGJELÉSBŐL

A dualizmus mellett szóló másik érv a saját elménkkel kapcsolatos kiváltságos önvizsgálati tudatosságunkból indul ki.

Elmének sajátos jellemzője, hogy rendelkezik reflexiós képességgel: gondolkodni tudunk saját gondolatainkról, saját mentális állapotainkról. Sőt, *egyedülálló* és *kiváltságos* hozzáféréssel rendelkezünk saját mentális állapotaink kontenciáihoz. Ez a hozzáférés annyiban egyedülálló, hogy csak én és csakis én vagyok beavatott a mentális életembe. Kiváltságos, mivel - ellentétben a világegyetemben minden máshoz való hozzáféréssel - a saját mentális életemhez való hozzáférésem *közvetlen*, és nem az érzékeim közvetítik.

Tekintettel az elméknek erre a közvetlen önvizsgálatra való képességére és alkalmasságára, hajlamosak lehetünk arra, hogy különbséget tegyünk az elmék és a fizikai tárgyak között, a következőképpen.

Amikor önvizsgálatot tartok - amikor a mentális életemre reflektálok, és a mentális állapotaim tartalmát vizsgálom -, nem *úgy tűnik* számomra, hogy a mentális életem eseményei fizikai események lennének. A jégkrémre való gondolkodásom számomra csak annak tűnik - jégkrémre való gondolkodásnak. Egyáltalán nem tűnik úgy, hogy ez egy elektrokémiai kisülés lenne az agyamban. Egyáltalán nem tűnik úgy, mintha *bármilyen* fizikai esemény lenne. Tehát a fagyalatra való gondolkodásomnak *nem fizikai* eseménynek kell lennie - a mentalitásnak nem fizikai jelenségnek kell lennie, ebben az esetben pedig legalábbis a szubsztancia dualizmus mellett kötelezzük el magunkat.

Az előző bekezdésben bemutatott, *a belső megjelenésből származó érv* meglehetősen gyenge kapcsolatot teremt a premisszái - amelyek a dolgok számunkra való *megjelenésével* foglalkoznak - és a következtetése között, amely azt állítja, hogy a dolgok *létezése* összhangban van a dolgok *látszatával*. Az introspektív megjelenésből kiinduló évrre a nyilvánvaló válasz tehát az, hogy rámutatunk arra, hogy nincs szükségszerű kapcsolat a dolgok számunkra való megjelenése és a dolgok tényleges volta között. Valójában nagyon gyakran megtéveszt bennünket az, ahogyan a dolgok látszanak: egy meleg szellő egyáltalán nem *tűnik úgy*, mintha milliárdnyi molekula mozgási energiája lenne, és az elektromosság sem *tűnik úgy*, mintha elektronok áramlása lenne.

A dolgok *látszata* azonban semmiképpen sem tesz különbséget egyrészt egy meleg szellő és több millió molekula mozgási energiája, másrészt az elektromosság és az elektronok áramlása között. Hasonlóképpen, az, ahogyan az elménk *tűnik* számunkra, nem támaszkodhatunk a mentalitás nem-fizikai voltának megállapítására.

A látszatok tévedhetetlensége természetesen nem jelenti azt, hogy az emberiség *nem* lehet nem-fizikai. Csupán azt mutatja, hogy az introspektív megjelenésből kiinduló érv a mentalitás nem-fizikai volta mellett szóló érvként megbukik, mivel nem képes megteremteni a szükséges szükségszerű kapcsolatot a premisszák igazsága és a következtetés igazsága között.

2.3.3 A LÉNYEGES TULAJDONSÁGOKRAVONATKOZÓ ÉRV

Az introspektív megjelenésből származó érv arra a tényre hivatkozott, hogy az elmének vannak bizonyos alapvető tulajdonságai, amelyek a közönséges fizikai tárgyaktól hiányoznak, és fordítva. Egyrészt az elmék rendelkeznek a közvetlen önvizsgálat és a reflektív tudatosság olyan alapvető képességével, amellyel a közönséges fizikai tárgyak nem rendelkeznek. Az elmék és a fizikai tárgyak azonban más lényeges tulajdonságokban is különböznek egymástól.

A közönséges fizikai tárgyak lényegében nyilvánosan hozzáférhetőek (bárki megfigyelhet egy széket), míg az elmék nem (csak én tudom közvetlenül megfigyelni az elmémet). A közönséges fizikai tárgyak is lényegében kiterjedtek a térben - tömegük, alakjuk, helyük és egyéb térbeli tulajdonságaik vannak. Az elmék viszont lényegében gondolkodó dolgok: pusztán azért, hogy elmék, nem rendelkeznek térbeli tulajdonságokkal, ahogyan egy székek pusztán azért, hogy székek, térbeli tulajdonságaik vannak. Az elméknek pusztán az elme mivoltukból adódóan csak a gondolkodási képességükhöz tartozó tulajdonságaik vannak. Az elméknek nem lényeges tulajdonsága, hogy térben kiterjedtek, ahogyan a közönséges fizikai tárgyakkal, például a székeknek lényeges tulajdonsága, hogy térben kiterjedtek.

Tekintettel az elme és a közönséges fizikai tárgyak közötti alapvető tulajdonságok e radikális eltérésére, tehát a *természetben is különbséget kell tenni*.

- az elméknek a fizikai tárgyaktól eltérő dolgoknak kell lenniük. Ezért nem-fizikai entitásoknak kell lenniük.

Bár a *lényegi tulajdonságokból kiinduló érv, amelyet az* imént elpróbáltunk, kezdetben meggyőzőnek tűnik, egy kis gondolkodás eloszlatja erejét. Sok mindent mondhatnánk a lényegi tulajdonságok metafizikájáról, amelyre az érvelés épül. Jelen célokra azonban elég, ha felismerjük, hogy a természetbeni megkülönböztetés *nem* egyenlő a metafizikailag lényegi megkülönböztetéssel.

Más szóval, ugyanazt a stratégiát alkalmazhatjuk, mint amit az introspektív tudatosságról szóló érvelés hatástalanításakor alkalmaztunk - elismerjük a premisszák igazságát, de rámutatunk, hogy ez nem teremt szükségszerű kapcsolatot a következtetés igazságával. Adott, hogy az elmék teljesen különböznek bármi

10

mástól, amit a világegyetemben ismerünk. Ez azonban nem jelenti azt, hogy az elmék *más anyagból lennének*, mint az elmék.

minden mást a világegyetemben - vagyis hogy az elmék *nem fizikai* anyagból állnak.

Mint korábban, a lényegi tulajdonságokból kiinduló érv kudarca nem jelenti azt, hogy az elmék *nem lehetnek* nem-fizikaiak - csupán azt mutatja, hogy az elmék és a kanonikus fizikai tárgyak közötti (lényegi vagy egyéb) tulajdonságok radikális megkülönböztetése nem elegendő annak megállapításához, hogy az elmék nem-fizikaiak.

2.4 ÉRVEK A KARTÉZIÁNUS DUALIZMUS ELLEN

Most három érvet láttunk a dualizmus mellett. Az introspektív megjelenés érve és az esszenciális tulajdonságok érve mindkettő a test-lélek dualizmus széles körű megalapozására törekszik. A józan ész bizonyos intuícióival párosulva, amelyek az elme efficacy-jára vonatkoznak a testben bekövetkező változások előidézésében és fordítva, ezek a karteziánus dualizmust támogató érvekké válnak. A vallási érv különösen a karteziánus dualizmust igyekszik igazolni, mivel az elme és a test közötti kölcsönhatás az érvelés szempontjából alapvető fontosságú.

Már most aggódnunk kellene az elméletért, hiszen aktívan kerestük a legerősebb érveket mellette, és felfedtük, hogy egyiknek sem sikerül megalapozni a következtetéseit. A karteziánus dualista számára azonban még aggasztóbbak a következő negatív érvek.

2.4.1 A MÁS ELMÉK PROBLÉMÁJA

A karteziánus dualizmussal szembeni első ellenvetés, amelyet megvizsgálunk, a nézet egy problematikus következményét jelöli ki.

Eléggé képzetek lettünk a fizikai világegyetem vizsgálatában, és ehhez mindenféle módszer és berendezés a rendelkezésünkre áll. A nem-fizikai világegyetem vizsgálatában azonban tanácstalanok vagyunk.

Ha az elmék anyagtalanok, akkor nyilvánvalóan nem vizsgálhatók az ismert empirikus módszerekkel. Ez nem csak azt jelenti, hogy az elmék kívül esnek a tudomány hatókörén, hanem azt is, hogy nem lehet tudni, hogy *más embereknek* van-e elméjük vagy sem. Amennyire a karteziánus dualista képes megítélni, lehet, hogy neki van az egyetlen elméje a világegyetemben - az összes többi emberi test lehet, hogy csak egy agyatlan automata.

Ez egy *ismeretelméleti* aggodalom - az azzal kapcsolatos aggodalom, hogy mit tudhatunk -, amely együtt jár az elme tudományának ~~hívó~~ kapcsolatos *módszertani* aggodalommal. Az ellenvetés azonban nem leküzdhetetlen. A karteziánus dualista segíthet magának egyfajta választ adni mindezekre az aggályokra.

A tudományos módszertannal kapcsolatban rámutathatna, hogy a tudomány számára nem ismeretlen, hogy megfigyelhetetlen entitásokat tételezzen fel és vizsgáljon meg megfigyelhető következményeik vizsgálatával. Tehát, miközben fenn kell tartania, hogy az elmék egyszerűen nem alkalmasak a *közvetlen* empirikus vizsgálatra, reménykedhet abban, hogy a mentalitásnak lesznek megfigyelhető következményei, amelyeket meg lehet vizsgálni, és ezáltal a tudomány *közvetett módon* hozzáférhet az elmékhez.

Ami a mások elméjéről való mindennapi tudásunkat illeti, a karteziánus dualista a saját mentális életével és annak a tapasztalat és a viselkedés közvetítésében játszott szerepével analóg módon érvelhet. Feltételezhetően - mondhatná - valóban úgy gondolja, hogy másoknak is van elméjük (meglehetősen diffikult a világban anélkül közlekedni, hogy ebből a feltételezésből indulnánk ki). Akkor miért gondolja ezt? Feltehetően azért, mert megfigyelte, hogy a legjobb magyarázat arra, ahogyan más emberek viselkednek, az, hogy mentális állapotokat tulajdonítunk nekik.

Más szóval, tudod, hogy *ha* bizonyos tapasztalatokkal rendelkezel, az bizonyos *hiedelmekhez* és *vágyakhoz* (mentális állapotokhoz) vezet, amelyek bizonyos helyzetekben arra készítetnek, hogy bizonyos módon viselkedj. Megfigyeltél továbbá más embereket, akik éppen ilyen helyzetekben éppen ilyen módon viselkednek, és következőképpen feltételezed, hogy ők is osztoznak bizonyos hiedelmeidben és vágyaidban (például abban a hiedelemben, hogy ebédidő van, és hogy enni szeretnél), és felismered, hogy ezek a mentális állapotok fontos magyarázó szerepet játszanak viselkedésük megértésében.

A más elmék problémájára adott válasz a *legjobb magyarázatra való következtetésre* hivatkozik: a legjobb magyarázat arra, ahogyan más emberi testek mozognak a világban, ha olyan mentális állapotokat tulajdonítunk nekik, amilyenekről *tudom*, hogy nekem is vannak. Nyilvánvalónak kell azonban lennie, hogy bár ez a válasz bizonyítja, hogy hasznos az *a feltételezés*, hogy más embereknek van elméjük (a feltételezés hasznos előrejelző képességeket biztosít), de semmiképpen sem *bizonyítja*, hogy van. A dualista számára a más elmék problémája továbbra is fennáll.

2.4.2 OCKHAM'S RAZOR

William of Ockham középkori filozófus és jelentős logikus volt a XIV. század elején. Bizonyára hallottál már Ockham borotvájának egy elterjedt elferdítését, amely valahogy így hangzik: "a legegyszerűbb magyarázat gyakran a legjobb". Helyesen értelmezve Ockham borotvája az elméletalkotás módszertani korlátjaként szolgál.

Ockham borotvájának legpontosabb megfogalmazása a

metafizika területén az, hogy "ne terjeszd ki az ontológiádat a szükségszerűségeken túlra". Másképpen úgy is fogalmazhatnánk, hogy nem szabad többé posztulálnunk

számos aggasztó ismeretelméleti aggályt vet fel.

Úgy tűnik, nincs világi mód arra, hogy a karteziánus dualizmust megmentsük ettől az ellenvetéstől. A dualizmusnak azonban olyan módosított formáit is előállíthatjuk, amelyek visszalépnek a kölcsönhatás iránti elkötelezettségtől.

2.5 EGYÉB DUALIZMUSOK

Emlékezzünk vissza a 2.2. szakaszból a [D1]-[D4] négy tételre, amelyek a karteziánus dualizmust jellemzik. Az egyik módja annak, hogy a karteziánus dualizmus alapvető ontológiai intuitív elképzeléseit visszaszerezzük a kölcsönhatás problémájának elmarasztaló kritikájától, az, hogy feladjuk a [D3] és [D4] tételek iránti elkötelezettséget, így a fizikai test és a nem-fizikai elme közötti kapcsolatról szóló beszámolót kell megtalálnunk. Ez a stratégia a párhuzamosság és az occasionalizmus néven ismert teista dualista elméletekhez vezet.

Egy másik lehetséges stratégia az, hogy csak a [D4]-t adjuk fel, és fenntartjuk a [D1]-[D3] tételek iránti elkötelezettséget. Ehhez ismét szükség van az elme és a test közötti kapcsolat sajátos bemutatására, amelyet az epifenomenalizmus néven ismert elmélet szolgáltat. Foglalkozunk ezekkel a többi dualista elmélettel sorban.

2.5.1 PARALLELIZMUS

A párhuzamosság egy teista dualista elmélet. A parallelista fenntartja az elme és a test ontológiai függetlenségét, de tagadja, hogy kauzális kölcsönhatásban állnának egymással. Ha egyszer tagadjuk az any test és az immateriális elme közötti kauzális kölcsönhatást, akkor a látszólagos kölcsönhatást valamilyen más módon kell megmagyaráznunk. A parallelista egy mindentudó, mindenható lényre hivatkozik, hogy megmagyarázza a kapcsolatot.

A párhuzamfelelős szerint, amikor ez a mindentudó, mindenható lény (a továbbiakban egyszerűen "Isten") megteremtette a testek fizikai világegyetemét és az elmék nem-fizikai világegyetemét, úgy alakította ki a dolgokat, hogy bár a fizikai világegyetemben az események sorrendje ok-okozati szempontból teljesen független a nem-fizikai világegyetem eseménysorától, a két sorrend mégis tökéletes harmóniában van. Ez az előre meghatározott harmónia magyarázza a mentális és a fizikai életem közötti összefüggést. Valahányszor a kezemet a forró tűzhelyre teszem, a fájdalom fájdalmas voltának mentális élményét élem át, mert Isten eredetileg úgy állította be a dolgokat, hogy ez így legyen. Valahányszor a karom megmozdításának vágyának mentális élménye ér, a karom felemelkedésének fizikai eseménye azért következik be, mert Isten a dolgokat eredetileg úgy állította be, hogy ez így legyen. És így tovább.

Bár a párhuzamosság megkerüli a kölcsönhatás problémáját azáltal, hogy tagadja az anyagi testek és az immateriális elmék közötti *oksági* kapcsolatot, még mindig a többi ellenvetéssel is küszködik, amelyek a

A karteziánus dualizmus. Még mindig megragadunk a más elmékhez való hozzáférés hiányának ismeretelméleti diffikultúrájában, és még mindig az Ockham borotvája ellen vétünk. Ráadásul az elmélet nem tűnik se többnek, se kevésbé hihetőnek, mint az általunk javasolt karteziánus teista válasz a kölcsönhatás problémájára.

2.5.2 OCCASIONALISM

A párhuzamossághoz hasonlóan az occasionalizmus is egy teista dualista elmélet, amely tagadja az anyagi test és az anyagtalán elme közötti kölcsönhatást, és Istenre hivatkozik a kettő közötti kapcsolat megmagyarázására.

Az egyetlen különbség a párhuzamosság és az alkalizmus között valójában az, hogy míg az előbbi szerint Isten a fizikai események és a nem fizikai események sorozatát előre meghatározott harmóniában hozta létre, addig az alkalista szerint Isten ott és úgy lép közbe, ahol és ahogyan szükséges, hogy a két sorozat harmóniáját fenntartsa.

Az alkalista szerint tehát valahányszor a kezemet a forró tűzhelyre teszem, a fájdalom fájdalmas voltának mentális tapasztalatát élem át, mert Isten beavatkozik, hogy ez így legyen. Valahányszor az a mentális tapasztalatom, hogy meg akarom mozdítani a karomat, a karom felemelkedésének fizikai eseménye azért következik be, mert Isten beavatkozik, hogy ez így legyen. És így tovább.

A párhuzamosságtól való eltérést motiváló sajátos doktrinális megfontolások itt nem fontosak. A mi céljainkhoz elég azt megállapítani, hogy az alkalmiasság ugyanazokat az előnyöket élvezi és ugyanazokat a kritikákat éri, mint a párhuzamosság.

2.5.3 EPIFENOMENALIZMUS

A dualista nézetnél lényegesen meggyőzőbb az epifenomenalista nézet. Valójában, ellentétben az általunk tárgyalt többi dualista nézettel, számos epifenomenalistával találkozhatunk a kortárs kognitív tudományban.

Az epifenomenalista fenntartja a [D1]-[D3] tételeket, csak a [D4]-et utasítja el. Fenntartja a mentális és a fizikai közötti ontológiai különbséget, és fenntartja az anyagi test és az immateriális elme közötti kauzális kapcsolatot *is, de csak az egyik irányban*. Azt állítja, hogy a fizikai állapotok mentális állapotokat eredményeznek, de *tagadja* a problematikus fordítottját.

Az epifenomenalista nem ütközik az interakció problémájába, *mivel* fenntartja a fizikai világegyetem kauzális zártságát. Az epifenomenalizmus szerint minden fizikai eseményt *teljes egészében* és *kizárólag* az azt megelőző fizikai események magyaráznak. Más

szóval, minden fizikai effektusnak van fizikai oka. Amellett, hogy a fizikai

effekták azonban úgy tartják, hogy egyes fizikai állapotok mentális állapotokat is eredményeznek.

Az epifenomenalista kép tehát egy olyan fizikai ok-okozati láncolatról szól, amely tartalmaz néhány fizikai állapotot, amelyek mentális állapotokat is eredményeznek. Ezeket a mentális állapotokat oksági szempontból ineffikusztikusnak tartják.

- nem *csinálnak* semmit. A mentális állapotok e nézet szerint a fizikai állapotok pusztá *epifé-* nomái. Fizikai állapotokkal magyarázhatók, de maguk nem okoznak sem fizikai állapotokat, sem további mentális állapotokat. Valószínűleg nem teljesen világos önöknek, hogy miért akarunk fenntartani egy olyan elméletet, amely a mentális állapotokat ontológiailag elkülönülve látja a hozzájuk tartozó fizikai állapotoktól, de mind a fizikai, mind a mentális birodalomban oksági szempontból ineffikatívnak. Ez valószínűleg sokkal világosabbá válik majd a kötet

későbbi részében, amikor a mentális állapotok *birtoklásának* vagy *meglétének* kiváltságos, első személyű tapasztalatát tárgyaljuk.

Egyelőre csak az a fontos, hogy megértsük az elmélet mechanikáját, és azt, hogy miben különbözik az ontológiai dualizmus más fajtáitól, amelyekkel eddig találkoztunk.

megvizsgálva.

2.6 ANOMÁLIS MONIZMUS

Egy utolsó elmélet, amelyet legalább meg kell említenünk, mielőtt lezárnánk ezt a fejezetet, az anomális monizmus, más néven a kettős aspektus elmélete vagy egyszerűen a tulajdon dualizmusa.

Az anomális monizmus szigorúan véve nem dualista elmélet az e fejezetben szereplő többi elmélet értelmében, mivel az anomális monista *nem* szubsztancia dualista. Az általuk képviselt dualizmus - már amilyen - a *tulajdonságok*, nem pedig a szubsztanciák dualizmusa.

Az anomális monista szerint nem létezik nem-fizikai *al-állás*. Vannak azonban, állítják, a fizikai anyagnak *redukálhatatlanul nem-fizikai tulajdonságai*. Más szóval, bizonyos fizikai állapotok *kettős aspektussal* rendelkeznek - egyszerre vannak rendes fizikai tulajdonságaik és bizonyos nem-fizikai tulajdonságaik, amelyek nem redukálhatók (magyarázhatók) fizikai tulajdonságaikra.

Az irreducibilisan nem fizikai tulajdonságok metafizikájának részletes megértése egy kis filozófiai szofisztikát igényel. Az érdeklődő olvasót arra bátorítjuk, hogy kövesse a további olvasmányokra vonatkozó javaslatokat, hogy kutatásait irányítani tudja.

Valószínűleg megint csak nem egyértelmű, hogy miért hajlamos valaki fenntartani az anomális monizmust. Ugyanaz az anyag, amely remélhetőleg teljes mértékben megvilágítja az epifenomenalizmus

mögött meghúzódó intuíciókat, valamennyire világossá teszi az anomális monizmus mögötti motivációkat is.

3. FEJEZET

VISELKEDÉS

A következő elmélet, amelyet megvizsgálunk, a filozófiai behaviorizmus. Mielőtt azonban ezt megtennénk, hasznos lesz egy rövid kitérőt tennünk a pszichológia történetébe.

Az egyik jó ok erre a kitérőre az, hogy az uralkodó szellemi légkör, amelyben a filozófiai behaviorizmus először kialakult, olyan volt, amelyben az empirikus pszichológia mint "az elme tudománya" még mindig nem állt meg. A pszichológusok és filozófusok még mindig nagyon is próbálták kitalálni, hogy mi is a pszichológia feladata, és összehangoltan törekedtek egy olyan szilárd filozófiai elmélet megfogalmazására, amelyben a mentalitás empirikus vizsgálatra alkalmas. A pszichológia egyidejű jelenlétének megértése a szellemi világszínpadon jelentős betekintést enged a filozófiai viselkedéskutatók motivációiba.

A másik jó ok a kitérőre az, hogy a "behaviorizmus" kifejezés a pszichológusok szájából egészen mást jelent, mint a filozófusokéból. Mivel itt lehetőség van a ~~felte~~ érdemes szigorúan elkülöníteni a "behaviorizmus" két értelmét.

A pszichológia messze a legfiatalabb teljes jogú tudományos diszciplína, mivel a filozófiából a legújabban vált ki. Csak a huszadik század elején vált el a pszichológia, és vált önálló akadémiai szakterületté. Mint ilyen, nem kell messzire visszamennünk a történelemben, hogy nyomon követhessük a pszichológia keletkezését és születését. A történet a tizenkilencedik századi Németországban kezdődik.

3.1 KORAI EMPIRIKUS PSZICHOLÓGIA

A pszichológia történetének tárgyalása itt meglehetősen felületes lesz. A mi céljaink érdekében arra szolgál, hogy meghatározzunk néhány kulcsfigurát és alapvető hozzájárulást, amelyek a pszichológia mint különálló tudományos szakterület megszületéséhez vezettek. Nagyjából és nagy vonalakban a következőket mondhatjuk el

a pszichológia gyermekkorát és korai gyermekkorát három szakaszra osztják, amelyeket a tudományág elődei által alkalmazott módszerek különböztetnek meg.

3.2 ÉLETTANI PSZICHOLÓGIA

Bár a "pszichológia megalapítója" címet másnak tartogatjuk, Gustav Fechnernek (1801-87) kell tulajdonítanunk a pszichológia empirikus hagyományának elindítását és az első kvantitatív pszichológiai törvény kidolgozását.

Fechner előtt az empirikus fiziológiának hosszú hagyománya volt, de a mentalitást mindig csak a priori vizsgálták, kísérletileg soha. Fechner kezdetben fiziológusnak készült, majd a lipcei fizika, később pedig a filozófia professzora lett.

Fechner felfedezte, hogy az érzékszervi bemenet intenzitásának *érzékelése* logaritmikusan arányos az inger abszolút nagyságával. Például a hangerő érzékelése logaritmikusan arányos a hanghullámok abszolút nagyságával. Talán észrevették, hogy a hangok hangosságát számszerűsítő decibel-skála logaritmikus skála. Fechner eredménye az érzékszervi modalitásokban robusztusan megnyilvánuló mennyiségi összefüggésnek bizonyult.

Nem szabad alábecsülnünk ennek az eredménynek a jelentőségét az elme empirikus tudományának lehetőségét és pszichofizikai módszertanát illetően. A szellemtörténetben először látjuk a fizikai jelenségek és a mentális jelenségek közötti megfigyelhető és mérhető kapcsolat azonosítását.

A korai pszichológia fiziológiai hagyományának másik jelentős alakja Hermann Helmholtz (1821-94). Fechnerhez hasonlóan Helmholtz is széleskörű tudományos érdeklődéssel rendelkezett. Kezdetben Potsdamban filozófiai és filológiai, majd Berlinben orvosi tanulmányokat folytatott. Tudományos pályafutása során a fiziológia tanszékét Königsbergben, az anatómia és fiziológia tanszékét Bonnban, a fiziológia tanszékét Heidelbergben és a fizika tanszékét Berlinben foglalta el. Elnöke volt a heidelbergi Fiziológiai Intézet és a berlini Fizikai Intézet létrehozásának is.

Amellett, hogy Helmholtz jelentős mértékben hozzájárult az élettani optikához - többek között feltalálta a szemmikroszkópot és a szemmétert -, és fontos egységesítő eredményeket szolgáltatott az elméleti fizikában, kísérleteket végzett az idegvezetéssel kapcsolatban is. Korai

A kísérletek célja az volt, hogy megmérjék az idegi impulzusok állati vétagokban való terjedésének idejét. Később ezt a kutatást kiterjesztette emberi alanyokra is, és ezzel bevezetett egy sokoldalú kísérleti technikát, amelyet a pszichológiában ma is széles körben alkalmaznak - az *inger-reakció* idő mérését.

Fechner és Helmholtz úttörő munkája megalapozta azt az ~~apt~~ amelyre mások az elme független tudományát építették. Fechner megmutatta, hogy empirikus módszereket lehet alkalmazni a mentalitás vizsgálatára, és a pszichofizikát a további kísérletek gyümölcsöző területének tekintette. Helmholtz volt az első, aki megmutatta, hogy az inger-reakció idők mérése termékeny módszertan lehet a születő pszichológia számára.

Jelentős érdekesség, hogy e korai fiziológiai pszichológusok úttörő munkája csak a két kutató széleskörű tudományos érdeklődésének és interdiszciplináris képzésének eredményeként vált lehetővé. Ez intő jel azok számára, akik túlságosan szűkre specializálódnának, valamint a tudományágak közötti elemzés és az interdiszciplináris együttműködés értékének egyértelmű megerősítése.

3.3 INTROSPEKCIONISTA PSZICHOLÓGIA

Wilhelm Wundt (1832-1920) az, aki egyértelműen megérdemli a "pszichológia megalapítója" címet. Wundt hozta létre az első pszichológiai laboratóriumot, 1881-ben megnyitotta az első pszichológiai folyóiratot - *Philosophische Studien* - és 1894-ben megalapította a Kísérleti Pszichológiai Intézetet Lipcsében. Ő írta az első pszichológiai tankönyvet is, és a világ minden tájáról érkező végzős hallgatók légióit irányította, akik a pszichológiai szakemberek első generációjává váltak.

Wundt pszichológiai strukturalista programja központi jelentőséget tulajdonított az introspekciónak mint módszertani technikának. A program célja a tudat elemzése volt annak érdekében, hogy azonosítani ~~hossz~~ annak alapelemeit és az azokat összekötő törvényszerűségeket. Ezt olyan gondosan megtervezett kísérletek segítségével valósította meg, amelyekben képzett megfigyelők önvizsgálatot végeztek mentális állapotukról, és beszámoltak megfigyeléseikről.

Az egyik kulcsfontosságú módszertani szempont itt a megfigyelők képzése. Wundt úgy vélte, hogy csak a megfelelően képzett megfigyelők képesek megfelelő gondossággal és odafigyeléssel megfigyelni, és megfigyeléseiket megfelelő módon, elemzésre alkalmas módon jelenteni.

További érdekesség, hogy a kísérleteket milyen szigorúan tervezték és végezték el. Wundt és tanítványai és kollégái

közreműködtek a kísérletezés számos, ma már szabványos kritériumának kidolgozásában, mint például a kísérleti helyszín nyilvánossága, az eredmények megismételhetősége és az a képesség, hogy bizonyos változókat állandó értéken tartsunk, míg másokat módosítunk.

Az introspekcionista hagyomány másik fontos kutatója Hermann Ebbinghaus (1850-1909) volt. Ebbinghaus 1890-ben megalapította Wundt folyóiratának riválisát, a *Zeitschrift für Psychologie und Physiologie der Sinnersorgane* című folyóiratot, és pszichológiai kutatólaboratóriumokat hozott létre Berlinben és Breslauban. Wundthoz hasonlóan ő is nagy hatású pszichológiai tankönyvet adott ki.

Míg Wundt és követői úgy gondolták, hogy a pszichológia hatókörének megfelelően azokra a pszichikai funkciókra kell korlátozódnia, amelyeket "alacsonyabb rendű" mentális funkcióknak nevezhetünk, és empirikus programjukat kizárólag a mentális képzetek vizsgálatára összpontosították, addig Ebbinghaus az emlékezet tanulmányozására szolgáló kísérletek kidolgozásával foglalkozott.

Annak érdekében, hogy az emberi emlékezet alapjául szolgáló mechanizmusokba a már ismert dolgok potenciálisan szennyező hatásaitól elszigetelt módon betekintést nyerjen, Ebbinghaus nagyon nagyszámú nonszensz szótaglistát - mássalhangzó-magánhangzó-mássalhangzó szegmenseket, amelyeknek a nyelvben nem volt jelentésük - állított össze. Ezután hozzálátott ahhoz, hogy ezeket a listákat memorizálja, és aprólékos eljárásokkal mérte, hogy mennyire képes felidézni ezeket a szótagokat.

Ez a kutatás további kvantitatív pszichológiai elveket szolgáltatott, amelyek között központi helyet foglal el az emlékezet exponenciális hanyatlása. Ebbinghaus felfedezte, hogy a nonszensz listák felidézésének képessége kezdetben nagyon gyorsan, de idővel egyre lassabban romlik. Az emlékezet exponenciális hanyatlásának egyik következménye a tanulók számára a korai megerősítés döntő fontossága. Ebbinghaus azt is kimutatta, hogy míg a listák kezdeti memorizálása gyors hanyatlásnak volt kitéve, a listák felidézési képességének hanyatlási sebessége az ismétlések számával arányosan lassult. Ennek a megállapításnak nyilvánvaló következményei vannak a vizsgálati technikákra nézve.

E két eredményt összevetve azt látjuk, hogy ha valaki azt akarja, hogy az anyagot nagy pontossággal tudja felidézni, akkor az anyagot az első bemutatás után nagyon hamar újra kell ismételni, majd egyre hosszabb időközök után meg kell erősíteni az anyagot. Például egy előadáson bemutatott anyag felidézéséhez célszerű az anyagot még aznap, majd néhány nappal később, majd egy héttel később, majd egy hónappal később, és így tovább. A tanulási

technikákra vonatkozó modern tanácsok legapróbb vizsgálata is ilyen ajánlásokat fog tartalmazni.

Az introspektionista hagyomány utolsó említésre méltó alakja Oswald Külpe (1862-1915). Külpe Wundt tanítványa volt Lipcsében és

később Würzburgban rivális pszichológiai iskolát alapított. A lipcsei és a würzburgi iskola között két fő vitapont volt. A két iskola nem értett egyet a pszichológia megfelelő feladatkörét illetően, és ez a nézeteltérés a kísérleti módszertannal kapcsolatos vitát is magával hozta.

Míg Wundt és tanítványai úgy vélték, hogy a pszichológia legitim alkalmazási területe kizárólag a mentális képzetek vizsgálatára korlátozódik, Külpe és követői úgy gondolták - ahogy Ebbinghaus is -, hogy a "magasabb rendű" kognitív funkciók, például az érvelés empirikus vizsgálata fontos szerepet játszik a pszichológiában. A lipcsei és a würzburgi iskola tehát lényegében a mentalitás különböző aspektusait vizsgálta.

A másik vitás pont a wundtíanusok és a külperianusok között a megfigyelők szerepét érintette az introspekcionista kísérletezésben.

Wundt, mint korábban hangsúlyoztuk, jelentős jelentőséget tulajdonított a megfigyelők képzésének, hogy képzett introspektorok legyenek. Külpe ezzel szemben kizárólag *képzetlen* megfigyelőket használt. A történelem ebben a tekintetben Külpe-nek adott igazat: a

mai pszichológiai kísérletek általában megkövetelik, hogy az alanyok ne csak pszichológiailag képzetlenek legyenek, de ne is tudják (és gyakran megtevesztik őket) annak az adott kísérletnek a céljait, amelyben részt vesznek. A korai pszichológia introspekcionista hagyományát az a meggyőződés egyesítette, hogy az önmegfigyelés - akár képzett megfigyelők (Wundt), akár képzetlen megfigyelők (Külpe), akár saját magunk (Ebbinghaus) által - a mentalitás vizsgálatának kulcsa. Bár az önmegfigyelést a modern pszichológiában még mindig használják bizonyos mértékig (jellemzően különböző más módszerekkel kombinálva), az introspekcionista hagyomány kihalt.

a huszadik század elején. Ennek több oka is van.

Először is, egyre nyilvánvalóbbá vált, hogy a szellemi élet nagy része egyszerűen átláthatatlan az önvizsgálat számára. Nem vizsgálhatom például a nyelvi produkciót és megértést irányító mechanizmusokat pusztán introspekcióval.

Másodszor, az önvizsgálat meglehetősen megbízhatatlan, függetlenül attól, hogy ki végzi az önvizsgálatot. Az emberek köztudottan rosszul ismerik fel saját mentális állapotukat; a képzetlen megfigyelők különösen rosszul. A képzett megfigyelők viszont hajlamosak arra, hogy a megfigyeléseket a vélt elvárásaiknak megfelelően alakítsák ki. Továbbá az önmegfigyelés *maga is* mentális folyamat, és ezért hatással van az önmegfigyelés tárgyát képező mentális folyamatokra. Ha például a dühödre reflektálsz, akkor valószínűleg vagy feloldódsz, és ezáltal kevésbé leszel dühös, vagy egyre hevesebbé válsz, és ezáltal még dühösebb leszel.

Végül, csak az önvizsgálatot végző ágens van beavatva önvizsgálatának közvetlen eredményeibe. Ha az önmegfigyelő alanyok nem értenek egyet, nincs mód arra, hogy egy harmadik fél eldöntse a megfigyelési vitákat. Tehát míg a kísérleti helyzet, amelyben az önmegfigyelés történik, kielégítheti a nyilvánosság követelményét, maga az önmegfigyelési folyamat nem.

3.4 PSZICHOLÓGIAI BEHAVIORIZMUS

A fiziológiai és introszpeci- cionista pszichológusok alapozó munkája mellett két további fontos történelmi előfeltétel vezetett a pszichológiai behaviorizmus kialakulásához. Az egyik ilyen történelmi előzmény a pozitívizmus befolyásos tizenkilencedik századi doktrínája volt. Az Auguste Comte (1798-1857) és Ernst Mach (1838-1916) által képviselt pozitívizmus reakció volt a spekulatív metafizikára és a teológiai feltételezésekre, amelyekről úgy vélték, hogy megfertőzték a filozófiát. A pozitívizmus hívei úgy vélték, hogy a legitim intellektuális kutatásnak, vagyis a "pozitív tudománynak" kizárólag a megfigyelhető dolgokkal kell foglalkoznia. Minden olyan tantételt, amely a megfigyelhetőségen túlmutató entitásokat vagy folyamatokat tételezett fel, a pozitívizmus jelzővel illettek. pejoratív "áltudomány".

A pszichológiai behaviorizmusra gyakorolt másik jelentős hatás Ivan Pavlov (1849-1936) munkássága volt. Pavlov orosz fiziológus volt, a *reflexívek* elméletének megalkotója. Pavlov úgy vélte, hogy a környezeti inger és a viselkedési reakció közötti kapcsolat e reflexívek segítségével magyarázható. Kétségtelenül hallottál már Pavlov kutyájáról. Pavlov mutatta ki először, hogy a környezetében lévő táplálékinger hatására a kutya gyomrában akkor is megindul az emésztőnedvek áramlása, ha a táplálék soha nem jutott el a gyomorba. Egy *veleszületett reflex ív* létezését tételezte fel, hogy megmagyarázza ezt a kapcsolatot.

A leghíresebb azonban az, hogy Pavlov kimutatta, hogy a reflexívek lehetnek *kondicionáltak* és *veleszületettek* is. Pavlov kutyája esetében egy csengő szólalt meg, valahányszor ételt vittek az állat közelébe. Végül a kutya már a csengő hangjára is nyáladzani kezdett. Ez a viselkedés Pavlov szerint egy *kondicionált reflex ívet* mutatott.

Tekintettel ezekre az egyidejű történelmi tényezőkre - a pszichológia fellendülése, a pozitívizmus virágzása és Pavlov reflexívek elméletének fejlődése - a pszichológia megérett a paradigmaváltásra. Ezt John Watson (1878-1958) amerikai pszichológus hozta el.

Watson erősen a pozitívizmus híve volt, és következőképpen elutasította az introspekció mint pszichológiai eszköz létjogosultságát, azzal érvelve, hogy ahhoz, hogy pozitív tudomány legyen, a pszichológiának csak a megfigyelhető viselkedéssel kellene foglalkoznia. Ezzel a pszichológiát - amelyet korábban az emberi tudat tudományának tekintettek - az emberi viselkedés tudományává **te**

Watson az emberi viselkedést irányító veleszületett reflexívek felfedezése érdekelte, és annak vizsgálata, hogy milyen körülmények között kondicionálhatók a reflexívek az emberben. E célból csecsemőkön és kisgyermeken végzett kísérleteket, amelyeket a mai etikai bizottságok biztosan nem hagynának jóvá.

A veleszületett reflexívek jelenlétének bizonyítására a kisbabáknál Watson kimutatta, hogy bizonyos környezeti körülmények, például hirtelen hangos zaj vagy a támasz hirtelen elvesztése (azaz a csecsemők leejtése) félelmi viselkedést vált ki a csecsemőkből. Egyesek talán kevésbé tartják ezeket az eredményeket figyelemre méltónak.

A kondicionált reflexívek vizsgálata során Watson és kollégája, Rosalie Rayner Albert B.-n, egy tizenegy hónapos csecsemőn kísérletezett. Albertnek egy fehér patkányt mutattak be. Amikor a kisfiú kíváncsian kinyújtotta a kezét, hogy megérintse az állatot, a kísérletezők egy acélrúddal a feje közelében hangos zajt keltettek, és ezzel félelmi viselkedést váltottak ki Albertből. A patkánynak a hangos zajjal együtt történő hét ilyen bemutatása után - amelyek közül öt egy héttel később történt, mint az első kettő - azt találták, hogy Albert a patkány bemutatásakor önmagában is félelmi ~~választ~~ mutat. Ezt a kondicionált választ később tartósnak és hasonló ingerekre is átvihetőnek találták. Albert ugyanezt a félelmi reakciót mutatta egy fehér nyúl, és kisebb mértékben egy kutya, egy szőrmebunda vagy egy Mikulás-maszk bemutatására is.

Watson után a pszichológiai behaviorizmus vált uralkodóvá, és nagyjából az 1960-as évek végéig uralkodott a pszichológiában. Ebben az időszakban vált szokássá, hogy ahol csak lehetett, patkányokon végeztek kísérleti munkát, és ezzel kezdetét vette a pszichológiában az a hagyomány, amelyet néha - kedves vagy pejoratív módon - "patkányokkal és statisztikákkal" foglalkozónak neveznek.

Watson után messze a legnagyobb hatású pszichológiai viselkedéskutató egy másik amerikai, Burrhus Frederic (B. F.) Skinner (1904-90) volt. Skinnert a reflexívek kondicionálásának leghatékonyabb eszközeinek meghatározása érdekelte. Feltalált egy ma "Skinner-doboznak" nevezett eszközt, amelybe patkányokat lehetett elhelyezni. Ez a doboz tartalmazott egy kart, amelyet a patkányok le tudtak nyomni, és amelyet úgy lehetett beállítani, hogy

működtetéskor táplálékgolyót juttasson ki. A működtetés
körülményeinek megváltoztatásával

Skinner képes volt ennek megfelelően módosítani a patkányok viselkedését.

A klasszikus pavlovi kondicionálástól eltérően ez a viselkedésmódosulás nem csak a viselkedést megelőző ingerektől (csengő, patkány és hangos zaj), hanem a viselkedést *követő* környezeti ingerektől is függött. Skinner ezt *operáns kondicionálásnak* nevezte el, és kidolgozott egy kapcsolódó elméletet arról, hogyan lehet a legjobban effektálni az operáns kondicionálást bizonyos viselkedések pozitív és negatív megerősítésének szabályozásával.

Skinner azzal érvelt, hogy az ilyen operáns kondicionálás széles körben alkalmazható társadalmi mérnöki technikaként. Azt javasolta, hogy a bűnözői magatartást jobb lenne viselkedésmódosító technikákkal kezelni, mint börtönbüntetéssel büntetni, és megjelentetett egy széles körben olvasott regényt, a *Walden Two-t*, amelyben felvázolta az operáns kondicionálással irányított, tervezett társadalomról alkotott utópisztikus elképzelését. Ahogy az várható volt, ezeket az elképzeléseket a dicséret és az elítélés keveréke fogadta.

Az 1960-as évek végére a pszichológiában a behaviorizmus népszerűsége csökkent az újonnan megjelenő kognitív pszichológia javára. A pszichológia behaviorista felfogásába vetett hit elvesztésének jó okai vannak.

Egyrészt világossá vált, hogy a pozitívizmus egyszerűen szólva hamis tanítás. A modern tudomány gyakran foglalkozik azzal, hogy megfigyelhetetlen entitásokról elméleteket alkosson. A nem megfigyelhető dolgok ilyen kezelése nem teszi például az elméleti fizikát "áltudományos" jellegűvé.

Ennél is fontosabb, hogy egyre világosabbá vált, hogy a mentalitás számos lényeges aspektusa egyszerűen nem kapcsolódik közvetlenül a megfigyelhető viselkedéshez. Úgy tűnik, hogy a memória, a következtetések levonásának képessége és a nyelvi megértés képessége mögött álló mechanizmusok nem feltétlenül korrelálnak semmilyen konkrét viselkedési formával. Mint ilyenek, ezek a kognitív funkciók nem vizsgálhatók a behaviorista hagyomány szerint.

Miközben az a helyzet, hogy az általunk vizsgált három pszichológiai hagyomány mindegyike a történelem homályába veszett, az is igaz, hogy a modern pszichológia mindháromból megőrzött bizonyos technikákat. Míg azonban e hagyományok mindegyike módszertanilag kifejezetten homogén volt, a modern pszichológia módszertanilag kifejezetten heterogén. A pszichológiában még mindig van helye a pszichofizikai kutatásnak, az önbevallás még mindig széles körben használt technika, és még mindig van szerepe a viselkedés megfigyelésének. Nincs azonban

többé helye annak a módszertani monizmusnak, amely a pszichológia e korai időszakát jellemezte.

3.5 FILOZÓFIAI BEHAVIORIZMUS

A pszichológiai behaviorizmus, mint láttuk, egy *módszertani* nézet - egy doktrína arról, hogy miként kell a pszichológiát végezni. A filozófiai behaviorizmus ezzel szemben egy *analitikus* nézet - egy érdemi elmélet arról, hogy mik a mentális állapotok. A továbbiakban, amikor egyszerűen "behaviorizmusra" hivatkozom, akkor a filozófiai változatra fogok utalni.

A filozófiai behaviorizmust részben a pozitívizmus motiválta, és az ezzel kapcsolatos azon törekvés, hogy olyan elméletet alkossanak, amely lehetővé teszi az empirikus pszichológiát. A viselkedésemélet a dualizmust sújtó súlyos elméleti nehézségekre is ~~egy~~ Gilbert Ryle - *Az elme fogalma* (1949) című művében - híres módon azzal vádolta a karteziánus dualistákat, hogy "szellemet a gépben" tételeznek fel. A viselkedéskutatók elvetették azt a dualista elképzelést, hogy a mentális állapot kifejezések valamilyen anyagtalán szubsztanciában zajló eseményeket és folyamatokat jelölnek.

A behavioristák szerint a nyelvünkben használt kifejezések, amelyeket mentális állapotok jelölésére használunk, valójában egyszerűen csak kényelmes kifejezések, amelyekkel komplex viselkedési formákra utalhatunk. Amikor például azt mondjuk, hogy Jon szerelmes, *valójában azt* értjük alatta, hogy Jon buta vigyorral az arcán járkál, hajlamos úgy beosztani az idejét, hogy minél többet lásson az érzelmi tárgyából, hajlamos rossz verseket írni stb. Amikor azt mondjuk, hogy Tillie-nek fáj a foga, *valójában azt* értjük alatta, hogy Tillie grimaszol és nyög, hajlamos összeszorítani az állkapcsát, hajlamos fogorvost keresni stb. A mentális állapotok attribúciói a behavioristák szerint a viselkedésfajták tulajdonképpeni attribúciói: a mentális állapotokra vonatkozó kifejezések *jelentése* megfelelően a viselkedés szempontjából van meghatározva.

Itt gyorsan tisztázzuk a filozófiai (analitikus) behaviorizmus és a pszichológiai (módszertani) behaviorizmus közötti kapcsolatot. Az előbbi nyilvánvalóan maga után vonja az utóbbit. Ha a mentális állapotok *éppen* viselkedésfajták, akkor a mentalitás vizsgálata *ipso facto* a viselkedésfajták vizsgálata. A fordítottja azonban nem érvényes. Nem igaz, hogy egy módszertani behavioristának analitikus behavioristának kell lennie. Gondolhatjuk úgy, hogy a pszichológiának az emberi viselkedés tudományának kell lennie, de mégis fenntarthatjuk, hogy a mentális állapotoknak a viselkedéstől független (de a viselkedés megfigyelésén keresztül vizsgálható) létezése van.

A filozófiai behaviorizmus, helyesen értelmezve, egy reduktív szemantikai tézis, amely szerint a mentális állapotok kifejezéseinek elemzése magában foglalja a viselkedésről való beszédre való

redukciót. Lényeges, hogy a viselkedésről való beszédet a viselkedési *diszpozíciók* szempontjából konkretizáljuk, nem pedig egyszerűen a viselkedésről való beszédet.

a viselkedés szempontjából. Elvégre lehet, hogy Tillie-nek fáj a foga, mégsem mutat semmilyen kapcsolódó viselkedést - talán fél a fogorvostól, és igyekszik elrejteni a fájdalmát, nehogy rákényszerítsék, hogy fogorvoshoz menjen. A viselkedéskutató ezt úgy veszi figyelembe, hogy a mentális állapotokat alapvetően diszpozicionális elemzéseket ad, mivel a diszpozíciók *gátolhatóak*. A példa esetében a viselkedéskutató azzal érvel, hogy Tillie fogfájása *éppen* a nyögésre és grimaszolásra való hajlamot, az állkapocs összeszorítására való hajlamot, a fogorvos felkeresésére való hajlamot stb. *jelenti*; azonban a fogásztól való féelme gátolja ezeket a diszpozíciókat, és ez magyarázza a manifeszt viselkedés hiányát.

A behaviorizmusnak kétségtelenül van néhány filozófiai előnye. Először is megoldja a más elmék problémáját, mivel lemond a megközelíthetetlen, anyagtalan mentális szubsztancia iránti elkötelezettségről. Megfelel Ockham borotvájának, mivel nem terjeszti ki ontológiánkat a magyarázó szükségességen túl, mivel a mentális állapotok többé nem rendelkeznek független létezéssel. Megfelel annak az elképzelésnek, hogy a mentalitás fokozatosan jön létre, mivel az összetettebb viselkedés képessége *éppen* az összetettebb mentális állapotok képessége. Meg tudja magyarázni az agy jelentőségét a mentális életben, mivel az idegrendszer a viselkedés (és így a mentalitás) fizikai oka. Végül pedig a dualizmussal szöges ellentétben világos módszertant biztosít a pszichológiai vizsgálatokhoz.

Sajnos e vonzó elméleti előnyök ellenére a behaviorizmusnak számos leküzdhetetlen ellenvetéssel kell szembenéznie.

3.6 A FILOZÓFIAI BEHAVIORIZMUS SZEMBENI ELLENVETÉSEK

Kezdjük a három olyan ellenvetéssel, amelyek pusztán problematizálják a behaviorizmust, és áttérünk a további három olyan ellenvetésre, amelyek a behavioristák számára leküzdhetetlenek.

Az első dolog, amit meg kell jegyezni, hogy a diszpozíciós kritériumok a kapcsolódó mentális állapot hiányában is teljesülhetnek. Vegyük a színészek esetét. A fogfájós karaktert játszó színész diszpozíciója szerint nyög, grimaszol, összeszorítja az állkapcsát és így tovább, de nyilvánvalóan nem szenved fogfájástól.

Továbbá a diszpozíciós kritériumok nem teljesülhetnek a kapcsolódó mentális állapot jelenlétében. Vegyük a sztoikusok esetét. Egy sztoikus személynek lehet, hogy rettenetes fogfájása van, mégsem hajlandó semmilyen kapcsolódó viselkedést tanúsítani.

A diszpozíciók gátolhatóságáról szóló beszámoló célja, hogy valamennyire válaszoljon ez utóbbi ellenvetésre; azonban a viselkedéskutatónak még így is van némi diffikultusza.

magyarázkodni. A szereplők és a sztoikusok, valamint a diszpozíciók és a mentális állapotok szétválásának egymástól eltérő módjainak figyelembevétele problematizálja azt az elemzést, amely a mentális állapotokat a viselkedési diszpozíciókkal igyekszik *azonosítani*.

A behaviorizmussal szembeni harmadik, és még aggasztóbb ellenvetés arra összpontosít, hogy pontosan hogyan is kellene *teljes mértékben* meghatározni az egyes mentális állapotokkal azonosított diszpozíciók listáját. Röviden, az aggodalom az, hogy nincs mód arra, hogy a diszpozíciók felsorolásának végéről eltávolítsuk a "stb." vagy az "és így tovább" szavakat. A mentális állapotok kifejezéseinek viselkedéseméleti parafrázisai alapvetően befejezhetetlennek tűnnek, mivel nagyon sokféleképpen megnyilvánulhat egy mentális állapot a viselkedésben.

Vegyük például azt a kijelentést, hogy "Nicole szereti a gyermekeit". Milyen kapcsolódó diszpozíciókat kellene betöltenünk a parafrázált elemzésbe? Azt, hogy Nicole hajlamos arra, hogy gondoskodjon arról, hogy megfelelően táplálják őket, hogy melegen öltöztesse őket, hogy megölelje őket, hogy elmondja nekik, hogy szeretik őket, hogy nevelje őket, hogy aggódjon a jólétükért, hogy törődjön a gondjaikkal, hogy vigasztalja őket, ha megijednek, hogy megvédje őket a bajtól, hogy éjszakánként ébren fekvődjön és a jövőjükre gondoljon, és így tovább, és így tovább.

Az első három kifogás minden bizonnyal aggasztja a viselkedéskutatót, de némi ügyes lábmunkával talán megválaszolhatóak. A következő három ellenvetés azonban alkalmas arra, hogy még a legügyesebb lábú viselkedéskutatót is legyőzze.

Először is, a fájdalom *fáj*. Fájdalmasnak lenni lényegében a *fájdalmasság* első személyű, minőségi, privilegizált élményét jelenti. Pontosán ez a fájdalmas érzés az, ami jellemzi, hogy mit jelent fájdalmat érezni, és ami megkülönbözteti a *valódi* fájdalmas viselkedést a *színlelt* fájdalomtól. Röviden, van valami, amitől fájdalmasnak kell lenni. Így van ez más mentális állapotokkal is. Van valami, ami olyan *érzés*, *mint* szerelmesnek lenni, vagy dühösnek lenni, vagy izgatottnak lenni egy közelgő nyaralás miatt. A viselkedésemélet teljesen elmulasztja megragadni a mentális állapotok ezen alapvető szubjektív minőségi aspektusait.

Másodszorra a diszpozíciós elemzések rosszul szolgálnak magyarázatként. Amikor mentális állapotokat tulajdonítunk másoknak, azt jellemzően azért tesszük, hogy megmagyarázzuk és megértsük a viselkedésüket. Ha azonban a mentális állapotokat a viselkedési diszpozíciók szempontjából kell elemezni, akkor ezek az attribúciók körkörösek és nem informatívak.

Ha például megkérdezek egy fizikust, hogy miért török össze az

üveg, ha éles ütés éri, és a válasz az, hogy azért, mert az üveg törékeny, akkor az üveg törékenységeinek magyarázata a ~~válasz~~ magyarázataként szolgál. Ha ezután rákérdezek a ridegség jelentésére, és azt a választ kapom, hogy

a ridegség az a tulajdonsága az üvegnek, hogy éles ütés hatására hajlamos összetörni, semmit sem tanultam. Amit én kerestem, az a törekenység valamilyen magyarázata volt az anyag ~~há~~ tulajdonságai, nem pedig a diszpozíciós tulajdonságai szempontjából.

Hasonlóképpen, tegyük fel, hogy megfigyelem Wayne-t, amint besétál egy pizzériába, és megkérdezem Eloise-t, hogy mit csinál, és a válasz az, hogy Wayne éhes. Wayne éhségét a viselkedésének magyarázatául szánjuk. Ha ezután rákérdezek az éhség jelentésére, és azt a választ kapom, hogy az éhség *csupán* a *(többek között)* ételkereső viselkedésre való hajlamot *jelenti*, akkor semmit sem tudtam meg.

A behaviorizmussal szembeni utolsó ellenvetés egy belső kritika formájában jelenik meg: megmutatjuk, hogy a behaviorizmus a saját fényében megbukik.

A viselkedésemélet *reduktív* elmélet. Az elmélet lényege, hogy a mentális állapotokról való beszédet gondos elemzések során olyan elméleti kifejezésekkel helyettesíti, amelyek megfelelnek a megfigyelhetőség és a nyilvános ellenőrizhetőség pozitívista kritériumainak. A cél a mentális állapotokra való hivatkozás teljes *kiküszöbölése* azáltal, hogy a mentális állapotokról való beszédet a viselkedési diszpozíciókról való beszédre redukáljuk.

A mentális állapot kifejezések viselkedéseméleti parafrázisairól azonban kiderül, hogy *kikerülhetetlenül* utalnak a mentálisra. Ha azt mondjuk, hogy Tillie összeszorítja az állkapcsát, az többet jelent annál, mint hogy a karja állkapocsirányban felemelkedik. Ha azt mondjuk, hogy fogorvost keres, az többet jelent annál, minthogy fogorvos-irányba lendül. A "szorítás" és a "keresés" attribútumok inkább *agens* attribútumok. Ha ezeket mondjuk, azt mondjuk, hogy Tillie aktívan, ágenciálisan és *szándékosan* kapaszkodik, keresi és így tovább. Ez már mentális attribúció.

Az emberi viselkedés mindig is mentális jelenség. Lehetetlen olyan meggyőző diszpozíciós parafrázisokat felsorolni a mentális kifejezésekre, amelyek nem éppen olyan ágensekre utalnak, mint a "kapaszkodik", "keres", "szervez", "biztosít", "figyel" és így tovább. Mint ilyen, a mentálisról való beszéd kiküszöbölhetetlen, és a behaviorista nem elemezte ki a gépezetből a szellemet.

4. FEJEZET

NEUROANATOMIA

Most egy rövid kitérőt teszünk az elme filozófiai elméleteinek vizsgálatától, és a funkcionális neuroanatómia kezdetleges ismereteit fejlesztjük ki.

A neuroanatómiai bevezetés itt valóban nagyon felületes lesz. Ebben a fejezetben a céljaim meglehetősen szerények. Elsősorban szeretném bemutatni, hogy az agy egyes részei hogyan specializálódtak bizonyos funkciók feldolgozására. Különösen azt fogjuk látni, hogy a nyelvi képességünk erősen lokalizált, és egy meglehetősen rendkívüli neurobiológiai adaptáció szolgálja ki.

A második esetben a *neuronok* működésének alapvető megértését szeretném megadni. Ez jó szolgálatot tesz majd nekünk a könyv későbbi részében, amikor a mesterséges neurális hálózatokat vizsgáljuk. Összességében azt szeretném érzékeltetni, hogy az emberi agy milyen csodálatos és megdöbbentően összetett tárgy.

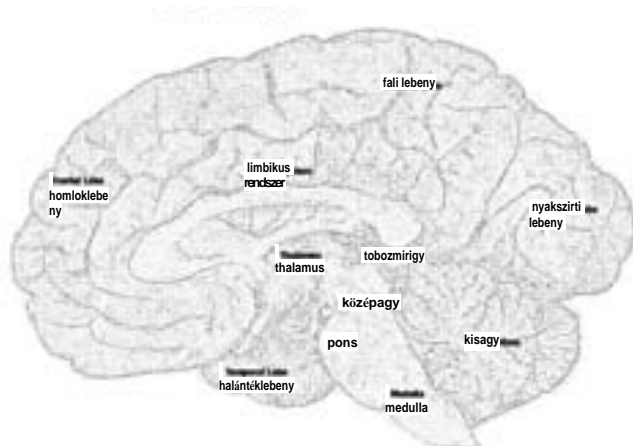
A makroneuroanatómia - az agy szabad szemmel is látható részei - leírásával kezdjük, majd áttérünk néhány alapvető mikroneuroanatómia leírására.

4.1 MAKRO-NEUROANATÓMIA

Az emberi központi idegrendszer nagyjából három területre osztható. A gerincvelő, az agytörzs és az agy többi része, beleértve az *agyféltekéket*, amelyek a *cerebrumot* alkotják.

A gerincvelő (*medulla spinalis*) érdekel minket a legkevésbé. Ez közvetíti a jeleket a tulajdonképpeni agy, valamint a szervek és az izmok között. A gerincvelő tetejével folytatódik az agytörzs, amely szintén három részre osztható.

Az alsó agytörzs, vagy *hátsó agy*, a *pons-t*, a *hosszúkás agyvelőt* és a *kisagyat* (nem összetévesztendő a *nagyagyvelővel*) tartalmazza. Az agyvelőről ismert, hogy részt vesz a szív működés és a légzés szabályozásában. A *pons* (híd) többnyire az információkat közvetíti.



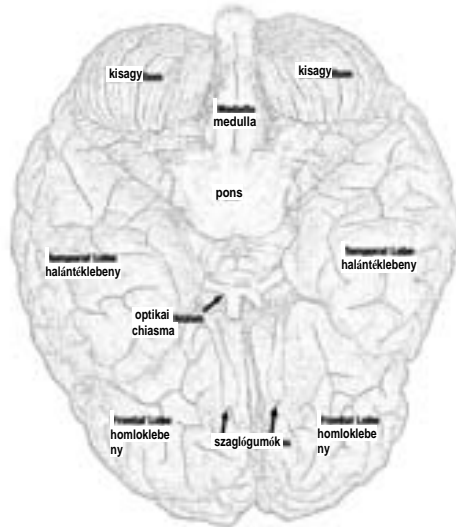
4.1. ábra Középső szagittális metszet.

az agyféltekék és a kisagy között, de a vestibuláris funkció (egyensúly) szabályozásában is szerepet játszik.

Az emberi *kisagy* (kisagy) rendkívül jellegzetes. Nagyon sűrűn vannak benne neuronok - sokkal sűrűbben, mint az agy többi részében -, és egy ilyen méretű idegrendszeri struktúrához képest meglehetősen szabályos a szerveződése. A kisagy könnyen felismerhető a felszínének nagyon finom redőzéséről, ami nagyobb felületet tesz lehetővé, és így áncos megjelenést kölcsönöz neki. Az emberi kisagy az emlősök agya között egyedülálló a maga összetettségében és bonyolult hajtogatásában.

A kisagy kapcsolatban áll a legtöbb elsődleges érzékszervi feldolgozó területtel és a legtöbb motoros neuronnal, és ismert, hogy részt vesz a finommotoros kontroll automatikus irányításában. Amikor például megtanulunk gépelni anélkül, hogy gondolkodnánk rajta, vagy gépjárművet vezetni anélkül, hogy gondolkodnánk rajta, a kisagyunk a különböző érzékszervi bemenetekre adott finom motoros válaszok sorozatának végrehajtására van programozva.

Az agytörzs következő része a *középgy*. A középgy összeköti a pons-t a felső agytörzssel. Ismert, hogy részt vesz a látással és a hallással kapcsolatos másodlagos feldolgozásban. Tartalmazza továbbá a *substantia nigra-t*, amely serkenti a *dopamin* nevű új transzmitter termelését, és amely szerepet játszik a finommotoros kontrollban. A Parkinson-kór, amelynek szufferoszeffei kontrollálhatatlan finom remegést tapasztalnak, a substantia nigra degeneratív állapota. Az agytörzs utolsó része - a *felső agytörzs* - tartalmazza a *talamuszt* és a *hipotalamuszt*, valamint a *tobozmirigy*et és az *agyalapi mirigy*et. A *talamusz*, amely folytonos a középső agyvelővel,



4.2. ábra Ép agy - alpnézet.

gyakran "érzékszervi kapuként" írják le. Az érzékszervi információkat (a szaglás kivételével) az érzékszervi készülékből a megfelelő elsődleges feldolgozó területre továbbítja. Több más struktúra mellett tartalmazza az *oldalsó geniculáris magot*, amely a szemekből érkező információk fogadására és az elsődleges vizuális feldolgozó kéregbe való továbbítására specializálódott, valamint a *mediális geniculáris magot*, amely a fülből érkező információk fogadására és az elsődleges hallásfeldolgozó kéregbe való továbbítására specializálódott.

A *tobzmirigy* és az *agyalapi mirigy* egyaránt *endokrin* mirigy - részt vesznek a neurotranszmitterek kiválasztásában. Az agyalapi mirigy szabályozza számos különböző neurotranszmitter, köztük az emberi növekedési hormon és az *oxitocin* felszabadulását. A tobzmirigy elsősorban a *melatonin* termelésével foglalkozik. Érdekes módon Descartes úgy vélte, hogy a tobzmirigy az anyagtalán lélekkel való kölcsönhatás helye.

A *hipotalamusz* irányítja a legtöbb *endokrin* mirigy szekrécióját, és ezáltal a testi funkciók széles körét szabályozza, beleértve a szomjúságot és az étvágyat, a hőmérsékletet, a szexuális és a cirkadián ciklusokat. Ez kimeríti az agytörzs struktúráit. Az agy többi része az *agyféltekékből*, a *limbikus rendszerből*, a *bazális és a bazális agyféltekékből* áll.

ganglionok és a *szaglógumók*.

A *szaglógumók* felelősek az elsődleges szaglási folyamatokért. Mint korábban említettük, a szaglás, vagyis a szaglás az egyetlen érzékszervi

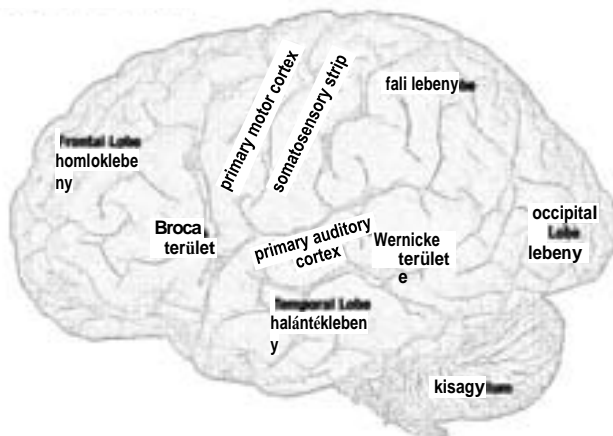
modalitás, amely nem a talamuszon keresztül irányul. Érdekes módon ez az egyetlen olyan érzékszervi modalitás, amely nem keresztirányú a féltekékben, ami azt jelenti, hogy - a többi érzékszervtől eltérően - a bal orrlyukból érkező információk a bal szaglógumóba, a jobb orrlyukból érkező információk pedig a jobb szaglógumóba jutnak.

A *bazális ganglionok* köztudottan a motoros processingben is részt vesznek. Szorosan kapcsolódnak a kisagyhoz és a középgyhoz. A bazális ganglionokkal kapcsolatos degeneratív állapotok, mint például a Huntington-kór (Chorea Huntington) esetén az érintettek kontrollálhatatlan testi mozgásokat tapasztalnak, amelyek jellemzően kifejezettebbek, mint a Parkinson-kórban szenvedőknél tapasztalt remegés.

A *limbikus rendszer*, amelyet néha *limbikus lebenyként is emlegetnek*, az agytörzs tetején helyezkedik el. Ez tartalmazza többek között a *hippokampuszt*, amelyről ismert, hogy részt vesz az emlékezet feldolgozásában, valamint az *amygdala-t*, amely az érzelmek tekintetében játszik szerepet. A limbikus rendszert néha az agyféltekék részeként írják le.

A legtöbb ember az *agyvelőt* alkotó agyféltekéket képzelel el, amikor az agyra gondol. Érdeemes megemlíteni, hogy az agy összes idegi struktúrája, nem csak az agyféltekék, a gerincvelő közepén hosszanti irányban húzott vonal mentén tükröződik. Az agyféltekéket egy vastag idegrostköteg, a *corpus callosum* köti össze.

A két agyfélteke a test érzékszervi és motoros funkcióinak egy-egy felének feldolgozásában vesz részt, és ezek a funkciók - a szaglás kivételével - keresztirányúak az agyban.



4.3. ábra Ép agy - bal oldali nézet.

féltekék. Tehát a bal félteke dolgozza fel a jobb oldali érzékszervekből származó információkat, és irányítja a test jobb oldali mozgását. A látás még sajátosabban keresztirányú. A szemek látómezejének jobb oldali részét a bal félteke, a szemek látómezejének bal oldali részét pedig a jobb félteke dolgozza fel.

Minden agyfélteke négy lebenyre oszlik: a *homloklebenyre*, a *fali lebenyre*, a *halántéklebenyre* és a *nyakszirti lebenyre*. Az agyvelő funkciói nem annyira jól ismertek, mint az eddig vizsgált, evolúciósan előbbre lévő agyterületek; vannak azonban bizonyos kanonikus kognitív funkciók, amelyek minden egyes lebenyhez kapcsolódnak, és vannak a lokalizált funkcionális specializáció fontos területei - nevezetesen az egyes érzékszervi modalitások elsődleges feldolgozó területei és a két beszédterület.

A *homloklebeny* a "magasabb" kognitív funkciók egész sorában játszik szerepet. Ezt néha lazán úgy fogalmazzák meg, mint ami a "tervezésért és előrejelzésért" vagy a "végrehajtó kontrollért" felelős. A homloklebenyben található az *elsődleges motoros kéreg*, amely, mint a neve is mutatja, a testi mozgások végrehajtásában játszik szerepet. Az agy egyik beszédterületét - a *Broca-területet* - is tartalmazza.

A Broca-terület, amely Paul Broca után kapta a nevét, az elsődleges motoros kéreg szomszédságában található. Az agy ezen részének károsodása *afázia* (nyelvi deficit) egy sajátos fajtáját, a *Broca-féle afáziát* eredményezi. A Broca-féle afáziára jellemző, hogy az érintett személy nem képes folyékony nyelvtani kifejezésekre, még akkor sem, ha a nyelvi megértés megmarad, és tudatában van a deficitjének. Ez egy különösen káros afázia lehet, mivel a suffererek küzdenek azért, hogy megértessék magukat, de csak kevés nyelvtani részecskét tartalmazó, vagy egyáltalán nem nyelvtani részecskét tartalmazó, szünetekkel tarkított megnyilatkozásokat tudnak produkálni.

A *parietális lebeny* vélhetően szerepet játszik az érzékszervi modalitásokból származó információk integrálásában. Ez tartalmazza az *elsődleges szenzoros kéregállományt*, más néven a *szomatoszenzoros szalagot*. A szomatoszenzoros szalag, akárcsak a szomszédos motoros szalag a homloklebenyben, *topológiailag* szervezett. Ez azt jelenti, hogy a szenzoros kéreg bizonyos részei a test bizonyos részeihez kapcsolódnak, és a kéreg nagyobb részei a test azon részeihez tartoznak, amelyek érzékenyebbek (több idegvégződéssel rendelkeznek). Így a szomatoszenzoros sáv nagy részei az ajkaknak, az ujjaknak és a nemi szerveknek vannak szentelve, de viszonylag kevés a kevésbé érzékeny területeknek.

Úgy gondolják, hogy a *halántéklebeny* bizonyos memóriafunkciókban játszik szerepet. Ez tartalmazza az *elsődleges hallókérget* és közvetlenül mellette az agy másik beszédterületét, a

A Wernicke-terület - amely Karl Wernicke-ről kapta a nevét - sérülésekor szintén egy disz- tinktív afázia - *Wernicke- afázia* - alakul ki. A Wernicke- afáziát folyékony, de értelmetlen beszéd jellemzi. Az érintettek jellemzően gyenge nyelvi megértést tanúsítanak, és kevéssé tudatosul bennük a hiányosságuk.

A nyelvi képességünket szolgáló rendkívüli biológiai adaptáció, amelyet korábban említettem, az *arcuata fasciculus*. Az arcuate fasciculus egy vastag idegrostsál, amely közvetlenül összeköti a Broca-területet a Wernicke-területtel. Az arcuate fasciculus károsodása *vezetési afáziát okoz*, amelynek egyik jellegzetes tünete, hogy nehezen ismételhünk vissza egy mondatot a beszélgetőpartnerünknek.

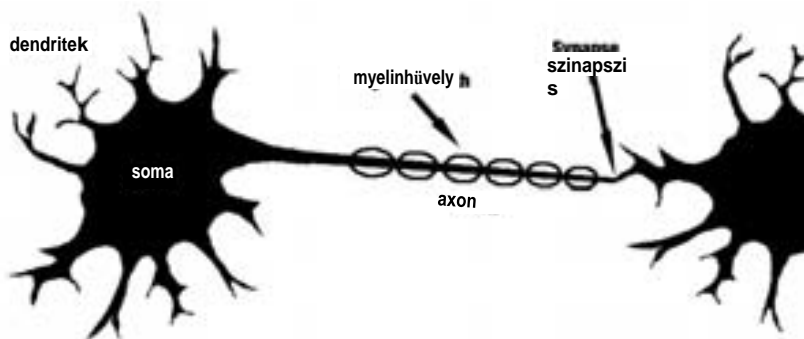
Végül a *nyakszirti lebeny* leginkább a vizuális információk feldolgozásával foglalkozik. Ez tartalmazza az *elsődleges látókérget* az agy leghátsó részén, ami megmagyarázza, hogy miért okozhat egy tarkóütés a fej hátsó részére azt, hogy az ember "csillagokat lát". A nyakszirti lebeny károsodása vakságot eredményezhet, még akkor is, ha a vizuális érzékszervek épek és működőképesek maradnak.

Bár az agyban a funkciók jelentős mértékben lokalizálódnak, bizonyos fokú *neurális plaszticitást* mutat, különösen a fiatalabb agyakban. Ez azt jelenti, hogy ha az agy egy bizonyos része károsodik, az agy más részei bizonyos mértékig képesek lehetnek átvenni annak funkcióját. Ez általában inkább az agyféltekékben megvalósuló funkciókat érinti. Az agytörzs károsodása általában visszafordíthatatlan, és jellemzően gyorsan halálhoz vezet, mivel ezek a területek létfontosságú funkciókat szabályoznak.

4.2 MIKRO-NEUROANATÓMIA

E fejezet utolsó célja a *neuronok* működésének rövid ismertetése. A neuronok olyan egyedi idegsejtek, amelyek elektromos impulzusokat vezetnek, és az agy igen nagyszámú ilyen sejtből áll.

Az agyban nagyjából tízmilliárd neuron található, amelyek mindegyike átlagosan körülbelül tízezer másik neuronhoz kapcsolódik. Ez *elképesztően* összetetté teszi az agyat. Képzeljük el, hogy veszünk egy India méretű országot - amelynek lakossága körülbelül egymilliárd fő -, és minden férfinak, nőnek és gyereknek ezer darab madzapot adunk, azzal az utasítással, hogy találjanak ezer különböző embert, aki a madzag másik végét fogja. Ha az egész ország így van összekötve, és minden ember ezer másik emberrel van összekötve a zsinórdarabokkal, akkor szorozzuk meg az egész rendszer bonyolultságát egy nagyságrenddel, és ennyire bonyolult az agyunk.



4.4. ábra Modell neuron.

A neuronoknak számos, egymástól meglehetősen különböző típusa létezik, de ez itt nem tartozik ránk. Egy paradigma neuron szerkezeti jellemzőit és működését fogjuk leírni.

Az idegsejteknek van egy *sejttestük* vagy *szomájuk*, amely a sejtmagot tartalmazza. Ez egy *axondombon* keresztül kapcsolódik az *axonhoz*, amely egy olyan pro- tuberancia, amely nagyjából egy méter hosszú is lehet. Ezeket az axonokat *mielinhévely* borítja, amely segíti az elektromos jelek gyorsabb áramlását és a szigetelést. Az axon végén *axonágak* vannak, amelyek *axonvégződése*kben végződnek.

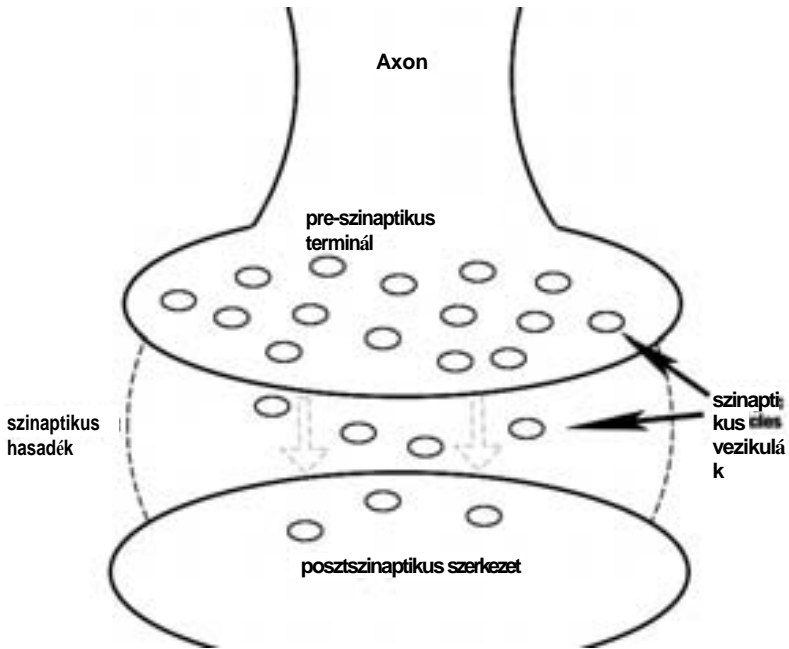
Az axonok *effere*ns kapcsolatok - jeleket szállítanak a szóma felől az axon mentén az axonvégzések felé. A bejövő, vagy *afferens* jelek a neuron *dendritjei* mentén a szóma felé haladnak. A dendritek *dendritfába* szerveződnek, és nagyon sok lehet belőlük.

Amikor egy axonvégződés egy dendrit közelébe kerül, *szinapszis* alakul ki (lásd a 4.5. ábrát). Ezek a *szinaptikus kapcsolatok* jeleket továbbítanak az egyik neuronból a másikba (szigorúan véve egy neuron a szomszédos neuron bármelyik részével szinaptikus kapcsolatot létesíthet, de most arra törekszünk, hogy a dolgok a lehető legegyszerűbbek legyenek).

A neuronok működése elektrokémiai természetű. Az elektromos jel az axon mentén a *preszinaptikus* axonvégződéshez áramlik, ahol kémiai jellé alakul át. Ezt a kémiai jelet aztán a neurotranszmitterek a *szinaptikus vezikulákban* átviszik a *szinaptikus hasadékon*.

Amint ezek a szinaptikus vezikulák elérik a *posztiszinaptikus* struktúrát, a neurotranszmitterek által kódolt kémiai jel visszaváltozik elektromos jellé, amely a posztiszinaptikus den- drit mentén terjed és a neuron testébe jut.

A neuron sejttestének van egy bizonyos elektromos *nyugalmi potenciálja*. Ahogy dendritjein keresztül különböző elektromos jeleket kap, a



4.5. ábra Modell szinapszis.

a sejt belseje és a sejt külseje közötti elektromos potenciál differenciája megnő. Amikor ez a potenciáldifferencia elég nagy, a szoma elektromos impulzust bocsát ki axonja mentén, és visszatér nyugalmi potenciáljára. Ez egyfajta egyszerűsítés, de a mi céljainknak megfelel.

Ha egy kicsit túlterheltnek érzed magad a technikai részletek és az új szakkifejezések miatt, ne bosszankodj. Ebből a fejezetből a neuronok működésével kapcsolatban csak a következőket kell magaddal vinned.

Az agyban nagyon sok idegsejt van, amelyek nagymértékben összekapcsolódnak egymással. Ezek a neuronok úgy működnek, hogy elektromos jeleket továbbítanak egymásnak. Ha egy idegsejt más idegsejtektől bejövő jeleket kap, akkor saját jelet küld.

Azoknak az olvasóknak, akiket érdekel, amit itt olvastak, és többet szeretnének megtudni, javasoljuk, hogy kövessék a további olvasásra vonatkozó javaslatokat.

AUSZTRÁL MATERIALIZMUS

Most, hogy legalább kezdetlegesen megértettük, milyen csodálatos dolog az emberi agy, itt az ideje, hogy megvizsgáljunk egy filozófiai elméletet, amely nagyon erős kapcsolatot feltételez az idegi és a mentális között.

Az ausztrál materializmus - így hívják, mivel fő képviselői ausztrál egyetemeken helyezkedtek el - egy olyan elmélet, amelynek sokféle neve van. Többféleképpen nevezik reduktív materializmusnak, identitáselméletnek, típusfizikalizmusnak és központi állam materializmusnak is, olyan okokból, amelyek a későbbiekben nyilvánvalóvá válnak.

Céljainkat szolgálja, ha az ausztrál materializmust egy másik elmélettel, az elme kauzális elméletével együtt fejlesztjük ki.

Amikor feltesszük a kérdést, hogy "mik a mentális állapotok?", kétféle választ adhatunk. Az egyik válasz a mentális állapotok fogalmi elemzését foglalja magában - egy beszámolót arról, hogy mit *értünk* "mentális állapotok" alatt. A másik fajta válasz lényegi azonosítást tartalmaz - annak megjelölésével, *hogy mely dolgok bizonyulnak* mentális állapotoknak.

A behaviorizmus esetében ez a kétféle válasz arra a ~~kérdésre~~ hogy mik a mentális állapotok, egy elméletben egyesült. Ennek az az oka, hogy a behavioristák a mentális állapotokkal kapcsolatban *eliminativisták*, és ezért nem hisznek abban, hogy létezik érdemi azonosítás. Szerintük a mentális állapotokról való beszéd valójában csak a viselkedési diszpozíciókról való beszéd - a behaviorizmus ontológiailag eliminatív és szemantikailag reduktív elmélet.

Most filozófiai kifinomultságot fogunk alkalmazni, és gondosan szétválasztjuk a mentális állapotok kérdésének kétféle megválaszolási módját. Az ausztrál materializmus a szubsztantív azonosítással fog szolgálni számunkra - azzal a beszámolóval, *hogy mely dolgokról derül* ki, hogy mentális állapotok. Arról azonban, *hogy mi az*, ami mentális állapot, az elme kauzális elmélete fog számot adni.

5.1 AZ ELME OKSÁGI ELMÉLETE

Az oksági elmélet kanonikus kifejtését David Armstrong adja 1968-ban megjelent *A materialista elmélet az elméről* című monográfiájában. Armstrong J. J. C. Smarttal és U. T. Place-szel együtt az ausztrál materializmushoz kapcsolódó három fő alak egyike.

Az oksági elmélet, mint már mondtuk, arra törekszik, hogy számot adjon *arról, hogy mi az*, hogy mentális állapot. Ez, ha úgy tetszik, nagyon hasonlít a mentális állapotok munkaköri leírásának megadásához. Egy munkaköri leírás nem határozza meg a szerep betöltőjének faját, korát vagy nemét. Csupán azt mondja el, hogy mik a vonatkozó feladatok - mit kell *tennie* valakinek ahhoz, hogy betöltse a szerepet. Így van ez az oksági elmélettel is - megmondja nekünk, hogy valaminek mit kell *tennie* ahhoz, hogy betöltse egy mentális állapot szerepét.

Kezdjük azzal, hogy elgondolkodunk azon a tényen, hogy nyelvünkben számos kifejezést az ok-okozati erőre való hivatkozással határozzuk meg. A "méreg" kifejezés paradigmátikus példa erre. A méreg lehet folyadék, szilárd anyag vagy gáz. A mérgek lehetnek színesek vagy színtelenek; lehetnek szagtalanok vagy jellegzetes szagúak, és így tovább. E tulajdonságok egyike sem releváns annak szempontjából, hogy a kérdéses anyagot megfelelően "mérregnek" nevezzük-e vagy sem. Egy anyagot az tesz méreggá, hogy milyen ok-okozati szerepe van abban, hogy az emberekben egészségkárosodást idézzen elő. Ha ezt definíciónak tekintjük, a következőket mondhatjuk: egy anyag akkor méreg, ha alkalmas arra, hogy az emberekben egészségkárosodást okozzon.

Az oksági elmélet központi tétele, hogy a mentális állapotok kifejezései éppen ilyen kifejezések. A kauzális teoretikus szerint a mentális állapot kifejezéseket a viselkedés tekintetében betöltött kauzális szerepük alapján határozzuk meg.

Ez lehetővé teszi az oksági teoretikus számára, hogy az egyes mentális állapotok meghatározására a következő sémát adja meg. Partikuláris mentális állapotban lenni olyan állapotot jelent, amelyet bizonyos ingerek *okozhatnak*, és amely *alkalmas arra, hogy bizonyos viselkedést váltson ki*.

A mentális állapotok meghatározó rendszertanának megalkotása tehát egyszerűen azon ingerek azonosításának kérdése, amelyek alkalmasak az egyes mentális állapotok kiváltására, valamint azon viselkedések azonosítása, amelyeket az egyes ingerek kiváltanak. Fájdalmat érezni például olyan állapotot jelent, amelyet *többek között* a kezem megégetése okozhat, és amely *többek között* a kezemnek a hőforrástól való eltávolítását és további, a kezemet megkülönböztető viselkedést okozhat.

Az oksági elmélet filozófiai elődjétől kezdve megőrzi azt az intuíciót, hogy a mentalitás és a viselkedés között döntő kapcsolat van. Nem teszi azonban a mentális állapotok és a viselkedési diszpozíciók közötti problematikus azonosítást. Egy mentális állapotban lenni a kauzális elmélet számára nem csupán azt jelenti, hogy bizonyos viselkedésre diszpozícióval rendelkezünk. Inkább azt jelenti, *hogy mentális állapotban lenni olyan állapotban van, ami*

alkalmasak arra, hogy bizonyos ok-okozati viszonyokban álljanak, amelyek az inger és a válasz között közvetítenek. Az oksági teoretikus nem eliminativista a mentális állapotok tekintetében.

A következő nyilvánvaló vizsgálati pont az, hogy pontosan meghatározzuk, mely állapotok alkalmasak arra, hogy rendelkezzenek ezekkel az oksági tulajdonságokkal. Más szóval, most, hogy a mentális állapotok szerepét a mentális állapotok fogalmi elemzésével jellemeztük, itt az ideje, hogy érdemi azonosítások révén lokalizáljuk e szerepek betöltőit.

Az oksági elmélet, ahogyan van, ontológiailag semleges - nem kötelez bennünket semmilyen konkrét ontológia mellett. Lehetünk például kauzális teoretikusok, de mégis fenntarthatjuk, hogy a mentális állapotok szerepét betöltő dolgok immateriálisak. Ehhez persze a kauzalitás szokatlan bemutatására lenne szükség, de már láttuk, hogy ez a dualisták számára problémát jelent.

Bár az oksági elmélet szigorúan véve ontológiailag semleges, a kauzalitásról való beszéd mégis elég nyilvánvaló utat nyit a mentális állapotok materiális azonosításához. Az ausztrál materializmus pontosan ezt kínálja.

5.2 AZ IDENTITÁSELMÉLET

Az ausztrál materializmus az 1950-es évek végén két nagy hatású tanulmány megjelenésével vált ismertté: *Place Is Consciousness a Brain Process?* (1956) és *Smart Sensations and Brain Processes* (1959) című munkája.

Az ausztrál materializmus szigorúan azonosítja a mentális állapotok és az idegi állapotok típusait. Más szóval, egy bizonyos típusú mentális állapotban lenni *csak annyit jelent, hogy egy bizonyos típusú neurális állapotban lenni.*

Ez a mentális állapotok elemzése, amelynek célja az *intertheoretikus redukció*. A mentális állapotok típusai az ausztrál materialisták szerint simán redukálódnak az idegi állapotok típusaira. A mentális állapotok tekintetében nem tartják az eliminativizmust, hanem inkább arra törekszenek, hogy mentális állapotokra vonatkozó kifejezéseink helyes referenseinek tudományos azonosítását végezzék el.

Az ausztrál materialista tehát nem hiszi, mint a behaviorista, hogy egyszerűen tévedünk, amikor a mentális állapot kifejezéseket úgy használjuk, mintha azok lényegi entitásokra vonatkoznának. Míg a behaviorista a mentális állapot kifejezéseket olyan kifejezésekhez tartotta hasonlóknak, mint a "boszorkányok" és a "flogiszon" - amelyekről a tudományos felfedezések előrehaladtával kiderül, hogy pusztán "népi" kifejezések -, addig az ausztrál materialista szerint a mentális állapot kifejezések olyan kifejezésekhez hasonlítanak, mint

a "villám", amelyeket tudományos elméleteinknek megfelelően fizikai jelenségekkel azonosítunk.

Ezen a ponton a különböző neveknek, amelyekkel az elméletet ismerik, már sokkal több értelme van. Ez egy reduktív materialista elmélet, amely típus-típus-azonosításokat tesz a mentális állapotok és bizonyos fizikai állapotok - nevezetesen a neurális állapotok, vagyis a központi idegrendszer állapotai - között - innen a "reduktív materializmus", "azonosságelmélet", "típusfizikalizmus" és "központi állapot materializmus" elnevezések.

Az elmélet egyik egyértelmű előnye, hogy megoldást kínál a más elmék problémájára. Meg tudjuk állapítani, hogy más embereknek valóban vannak-e mentális állapotai, egyszerűen az agyuk vizsgálatával. Egyfajta neurális állapottal rendelkezni, ezen elemzés szerint, egyszerűen egyfajta mentális állapottal rendelkezni, így más elmék könnyen azonosíthatóak és empirikusan megismerhetőek.

Az elmélet másik egyértelmű előnye, hogy tudományos módszertant biztosít a mentalitás vizsgálatához. Ha az elméről akarunk tudni, akkor az idegtudományt kell művelnünk. Különösen arra kell törekednünk, hogy meghatározzuk, hogy az idegi állapotok mely típusai milyen típusú mentális állapotokat kapnak.

Materialista - vagy fizikalista - elméletként az ausztrál materializmus megfelel Ockham borotvájának is. Az ausztrál materialista csak ~~fü~~ szubsztanciát fogad el ontológiájában. Legalábbis *qua* ausztrál materialista ez a helyzet - lehet, hogy más okai is vannak ontológiája kiterjesztésére, de ezek nem olyan okok lesznek, amelyek az elmeelméletére vonatkoznak.

A végső érv az elmélet és az elme oksági elméletének elméleti illeszkedésében rejlik. Amennyiben valaki úgy véli, hogy az oksági elmélet a mentális állapotok helyes elemzése, akkor előnyösnek találja, hogy az ausztrál materializmus olyan jelölteket ~~á~~ fel, amelyek pontosan azokkal az oksági erővel rendelkeznek, amelyeket a mentális állapotok jellegzetes meghatározójának tartanak.

Ha tehát az oksági elmélet fogalmi elemzését összekapcsoljuk az ausztrál materializmus tartalmi azonosításával, akkor a mentális állapotokról a következő beszámolót kapjuk. Egyfajta mentális állapotban lenni annyit jelent, mint egyfajta neurális állapotban lenni, amelyet bizonyos ingerek képesek kiváltani, és amely alkalmas arra, hogy bizonyos viselkedést okozzon.

Az ausztrál materializmus számos előnye ellenére, mint mindig, most is számos filozófiai ellenvetést tehetünk az elmélet ellen.

5.3 ÉRVEK AZ AUSZTRÁL MATERIALIZMUS ELLEN

Kezdjük néhány meglehetősen gyenge ellenvetéssel az ausztrál materializmussal szemben. Érvelhetünk azzal, hogy képesek

vagyunk mentális állapotaink önvizsgálatára, és amikor ezt tesszük, akkor tanulunk mentális állapotainkról. Mi

azonban semmit sem tudunk meg a neurofiziológiánkról az önvizsgálat révén, így a mentális állapotok nem lehetnek azonosak a neurális állapotokkal.

Erre az ellenvetésre egyértelmű a válasz. Egyenesen az ausztrál materialistával szemben *teszi fel a kérdést*. A "kérdésfelvetés" egy olyan kifejezés, amelyet egyre gyakrabban használnak, hogy valami olyasmit jelentsen, mint "amelynek fényében a kérdést fel kell tenni". Ez *nem a "begs the question"* filozófiai szakkifejezés. A kérdés felvetése a *petitio principii* tévedésébe ütközik. Az ember a kérdést a beszélgetőpartnerével szemben teszi fel, amikor azt kéri, hogy a vitatott tételt meg kell adni neki.

Ebben az esetben a vitatott tétel az, hogy a mentális állapotok típusazonosak az idegi állapotokkal. Az introspekcióból származó ellenvetés csak akkor vágja meg a filozófiai jeget, ha feltételezzük, hogy a mentális állapotok *nem* azonosak a neurális állapotokkal. Végül is, ha az ausztrál materialistának igaza van, és ez a típusazonosság fennáll, *akkor az* introspekció révén valóban megtudunk valamit az idegi állapotainkról, egyszerűen azért, hogy megismerjük a mentális állapotainkat (feltéve, hogy ezeket azonosnak tekintjük). Az introspekcióval kapcsolatos ellenvetés tehát egyszerűen csak felveti a kérdést, és nem valódi ellenvetés. Meglepő lehet, ha megtudjuk, hogy valóban introspektáljuk neurális állapotainkat, de a tudományos felfedezések gyakran meglepőek a korábbi népi elméletek fényében.

Az ausztrál materializmussal szemben számos további ellenvetés is van, amelyek szintén felvetik a kérdést. Ezek közül több a Leibniz-törvényre való hivatkozás formájában érkezik - az introspekcióból származó ellenvetés tulajdonképpen éppen ennek a formának egy példája.

A Leibniz-törvény - más néven *a megkülönböztethetetlenek azonossága* - azt állítja, hogy ha két dolognak minden és csakis ugyanazok a tulajdonságai, akkor azonosak. Az introspekcióból származó ellenvetés ezt igyekszik bevetni azzal érvelve, hogy mivel a mentális állapotok rendelkeznek egy olyan tulajdonsággal, amellyel az idegi állapotok nem, ezért nem lehetnek azonosak. Hasonló ellenvetéseket lehet felhozni különböző más tulajdonságokra hivatkozva, amelyekkel az ellenvetést tevő szerint a mentális állapotok rendelkeznek, az idegi állapotok pedig nem, vagy fordítva.

Az idegi állapotoknak például meghatározott tér-időbeli elhelyezkedésük van. Furcsának tűnik azonban az a feltételezés, hogy a jégkrémre való gondolkodás mentális állapotom három centiméterrel a jobb szemem mögött található.

Alternatívaként játszhatunk a mentális állapotok szemantikai tulajdonságaival is. Az a mentális állapotom, hogy azt hiszem, ma

szombat van, *szemantikai tartalommal* bír - *jelent* valamit. Szemantikai tartalmánál fogva alkalmas arra, hogy implikációs relációkban szerepeljen - például, *ha* azt hiszem, hogy ma szombat van, *akkor* azt hiszem, hogy holnap vasárnap lesz. Az idegi állapotoknak azonban sem szemantikai tartalma nincs, sem pedig implikációs relációkban való részvételre nem alkalmasak.

Számos más módon is problematizálhatnánk a Leibniz-törvényre való hivatkozást; a legdőntőbb válasz azonban az, ha egyszerűen megjegyezzük, hogy ezek az ellenvetések megint csak egyenesen az ausztrál materialistával szemben teszik fel a kérdést. Ha a mentális állapotok és a neurális állapotok valójában típusazonosak, akkor a mentális állapotoknak *valóban* van egy meghatározott tér-időbeli helyük - bármilyen meglepő is ez -, és a neurális állapotoknak valóban van szemantikai tartalmuk, így alkalmasak arra, hogy implikációs kapcsolatokban vegyenek részt, bármilyen meglepő is ez. Ha egyszerűen azt *állítjuk*, hogy ez nem így van, az csak a kérdés felvetése.

Van azonban egy meglehetősen határozott ellenvetés az ausztrál materializmussal szemben. Ez a többszörös megvalósíthatóságból fakadó ellenvetés.

Kiderült, hogy nem túl diffikult komoly filozófiai aggályokat vetni fel azzal az állítással kapcsolatban, hogy a mentális állapotok típusai azonosak az idegi állapotok típusaival. Először is, felmerül a kérdés, hogy miként kell értelmeznünk a "típus" fogalmát. Ha túl szűken értelmezzük, akkor azt kell mondanunk, hogy amikor egy csoportunk mindannyian jégkrémre vágyik, akkor mindannyian *pontosan ugyanabban* az idegi állapotban vagyunk. Ez nyilvánvalóan valószínűtlen, és a "típus" fogalmának a neurális állapotok némi variációját kell lehetővé tennie, hogy ezt figyelembe vegyünk. A döntő kérdés az, hogy *mekkora* a variáció mértéke. Ha a "típus" fogalmát túl tágan értelmezzük, akkor fennáll a veszélye, hogy elveszítjük azt az empirikus módszertani előnyt, amelyet a típusazonosság állítása biztosít.

Az ausztrál materialista magyarázati terhe itt az, hogy valamilyen módon számot adjon arról, hogy pontosan miben kell osztozniuk a neurális államoknak ahhoz, hogy azonos típusúnak minősüljenek. Sajnos az ausztrál materialista számára, még ha ez a magyarázati teher teljesíthető is, van egy további ellenvetés a típusazonossági állítással szemben, amely megválaszolhatatlan.

Vegyük például azt az esetet, amikor valaki idegkárosodást szenved, akár agyvérzés, akár valamilyen trauma, például gépjárműbaleset következtében. A károsodás után a beteg jellemzően elveszíti bizonyos mentális állapotok képességét. Lehet, hogy például már nem képes felismerni a házastársát, vagy nem képes többé bizonyos szavakat megérteni. A döntő fontosságú tény itt az, hogy az ilyen betegek nagyon gyakran visszanyerik sok elveszített mentális képességüket - *újra megtanulják* felismerni a házastársukat vagy megérteni az elvesztett fogalmakat. Visszanyerik az ilyen mentális állapotok képességét *annak ellenére*, hogy az idegrendszeri szubsztrát, amely eredetileg támogatta ezeket a funkciókat, visszavonhatatlanul károsodott.

A károsodás előtt az x mentális állapotban lenni azt jelentette, hogy az x neurális állapotban vagyunk, az elvesztett mentális funkciók újratanulása után azonban,

Az x mentális állapotban lenni azt jelenti, hogy egy *teljesen különböző* y neurális állapotban vagyunk. Ez a neurális plaszticitás - az agy azon képessége, hogy az agy egyes részei olyan funkciókat vesznek át, amelyeket egyébként az agy teljesen különböző részei látnak el - jól dokumentált. A mentális állapotok kimutatható többszörös megvalósíthatósága döntően cáfolja az ausztrál materialista által felállított típusazonosságot.

A mentális állapotok tehát nem csak többszörösen megvalósíthatók a különböző szubjektumok között oly módon, hogy az problematizálja a "típus" rovatot, hanem bizonyíthatóan többszörösen megvalósíthatók *egyazon szubjektumon belül is* oly módon, hogy az megcáfolja azt az állítást, hogy a mentális állapotok típusai azonosak a neurális állapotok típusaival.

Ezen kívül érdemes itt röviden megjegyezni, hogy az ausztrál materializmus előítéletes a nem emberi elmék lehetőségével szemben. Ha a mentális állapotokat az emberi központi idegrendszer állapotaival azonos típusúnak tekintjük, akkor például a kutyák és macskák számára nem lehetséges, hogy mentális állapotokkal rendelkezzenek. Egészen biztos vagyok azonban abban, hogy Linus macskánknak és Mia kutyánknak vannak mentális állapotai. Természetesen nem rendelkeznek olyan összetett mentális állapotokkal vagy kognitív képességekkel, mint az emberek.

– távolról sem - de a legkevésbé sem tűnik hihetőnek az az állítás, hogy nincsenek hiteik és vágyaik.

Az ausztrál materializmus eleve kizárja, hogy a mentális állapotok nem emberi biológiai szubsztrátumokban létrejöhessenek. További

– és ami a mi céljaink szempontjából döntő fontosságú - kizárja a mesterséges intelligencia lehetőségét. Ez önmagában véve nem nagy ellenvetés, de mindenképpen az elmélet nemkívánatos - és motiválatlan - következményének tekinthetjük.

A többszörös megvalósíthatósági kifogás fényében világos, hogy nem lehet továbbra is fenntartani a mentális állapotok és az idegi állapotok közötti típusazonosságot. Az elmélet egyik lehetséges módosítása az, hogy visszavonulunk a típus-token azonossághoz. Ez egyszerűen azt jelenti, hogy amikor valaki egy bizonyos típusú mentális állapotban van, akkor van egy hozzá tartozó token neurális állapot. Más szóval, egy mentális állapot típusában lenni csak annyit jelent, hogy *valamilyen* neurális állapotban vagyunk. A mentális állapotokat továbbra is azonosnak tekintjük a neurális állapotokkal, de a neurális állapotok egy adott típusát nem tekintjük a mentális állapotok egy adott típusának.

Bár ez a módosítás figyelembe veszi a mentális állapotok többszörös megvalósíthatóságát - mind az alanyok között, mind

ugyanazon az alanyon belül az időben -, mégis nagyon gyenge elméletet alkot. Az az elméletalkotó, aki a típus-jelölő azonosságot tételezi fel, már nem tesz olyan azonosítást, amely megkönnyíti az interteoretikus redukciót. Nem vizsgálhatjuk többé a mentalitást az idegtudományok segítségével, mivel nem jelent előnyt annak meghatározása, hogy mely mentális állapotok milyen neurális állapotokként érvényesülnek.

Ha a személyazonosság csak egy jelképes személyazonosság, akkor ez a meghatározás csak a vizsgálat tárgyát képező alanyra vonatkozik a vizsgálat időpontjában. Az eredmények nem univerzalizálhatók úgy, ahogyan a típusidentitást tételező teoretikus esetében.

A jelképes identitáselmélet - vagy jelképes fizikalizmus, ahogyan némelykor nevezik - tehát aligha érdemli meg, hogy szórakoztassuk. Az elmeelmélet filozófiai megfelelőségére vonatkozó egyik kívánalmunk az empirikus megfelelőség - ideális esetben az elméletünknek empirikus vizsgálatokat kellene irányítania.

Szerencsére most már rendelkezésünkre áll egy ilyen elmélet, és a következő fejezet célja ennek kidolgozása lesz. Ezt úgy fogjuk megtenni, hogy az elme kauzális elméletének központi intuícióját oly módon őrizzük meg, hogy lehetővé tegyük az ausztrál materializmus alapjául szolgáló motivációkat, de anélkül, hogy túlságosan elköteleznénk magunkat olyan módon, ahogyan azt a lényegi típus-típus azonosításuk teszi.

Mielőtt azonban ezt megtennénk, célunk érdekében megvizsgálunk egy nagyon jól ismert gondolat kísérletet, és megvizsgálunk egy lehetséges érvet, amelyet levonhatunk belőle.

5.4 AMIT MÁRIA NEM TUDOTT

A gondolat kísérlet fontos szerepet játszik az elmefilozófiában. Mivel ebben a kötetben először látunk ilyet, érdemes nagyon röviden kitérni a szerepükre.

A gondolat kísérletek célja az intuíciónk megalapozása azáltal, hogy arra kérnek minket, hogy képzeljünk el bizonyos logikailag lehetséges helyzeteket. Természetüknél fogva jellemzően vadul furcsa és valószínűtlen helyzeteket írnak le, és ez alól a következő gondolat kísérlet sem kivétel. Az azonban nem filo- szofikus munka, ha egyszerűen csak tiltakozunk a gondolat kísérleti helyzet bekövetkezésének fizikai lehetősége ellen, bár ez gyakori válasz, amikor először találkozunk velük. A gondolat kísérletekkel való filozófiai foglalkozás azt jelenti, hogy a leírt helyzet logikai következményeit azonosítjuk - azzal érvelünk, hogy ennek és ennek a helyzetnek így és így *kell lennie*, ha a helyzet bekövetkezik.

Ezt szem előtt tartva, nézzük meg Mary esetét, az empirikus ~~ut~~ par excellence. Mary hozzáférhet a befejezett fizikai elméletekhez - nem csak a jelenlegi legjobb elméleteinkhez, hanem a *befejezett* elméletekhez is. Mary már elég hosszú ideje szorgalmasan tanulmányozza ezeket az elméleteket, és eljutott arra a pontra, ahol már *minden* fizikai tény ismer. Különösen a színekről tud *mindent*, *amit* csak tudni lehet. Mindent tud a fény hullámhosszáról és a különböző felületek fényvisszaverő képességéről. Mindent tud az

emberi neurofiziológiáról is, így mindent tud a

az emberi érzékszervek és a vizuális képesség. Ismeri az *összes fizikai tényt, amit az emberi színélményről tudni lehet.*

Marynek azonban valójában soha nem volt színélménye.

Mostanáig egy olyan eszközt ültettek az agyába, amely megakadályozza, hogy színesben lásson - az egész világa eddig a fekete-fehér élmények világa volt. Jutalmul azonban, hogy végre mindent megtanult a fizikai világról, amit csak tudni lehet, ezt az eszközt távolról kikapcsolják, így Mary képes lesz színes élményeket átélni. Az első dolog, amit Mary a készülék kikapcsolása után lát, egy vörös rózsza. Mary olyan élményben részesül, amiben korábban még soha nem volt része - a piros szín látásának élményében. És *valami újat tanul.* Megtanul valamit, amit korábban nem tudott, annak ellenére, hogy minden ~~fényt~~ fényt ismert a világról. Megtanulja, *milyen* érzés látni a piros színt. E gondolat kísérlet eredetileg arra szolgált, hogy alátámasszon egy érvet *minden* olyan elmélet ellen, amely a mentalitást tisztán fizikai szempontból próbálja megmagyarázni. Az érvelés a következőképpen hangzik. Mary ismerte az összes *fizikai* tényt. Mégis Mária valami újat tanult, amikor színélményben volt része. Tehát a mentális életről többet kell tudni, mint amennyit az összes fizikai tény nyújt. Ezért egy tisztán a mentalitás fizikai beszámolója nem teljes beszámoló.

Ami kimaradt, az a mentális állapotról szóló beszámoló, állítják. A mentális állapotok ezen kiváltságos szubjektív minőségi aspektusait - *azt, hogy milyen az az élmény, hogy egy állapotban vagyunk* - *kvalitásoknak* nevezzük. A 3. fejezetben röviden érintettük a kvaliákat, amikor a fájdalom fájdalmas voltát mint a behaviorizmus ellenvetését tárgyaltuk. A fizikalizmus elleni adott érv világossá teszi ennek az ellenvetésnek a terjedelmét és lényegét.

Tagadhatatlan, hogy van valami, ami olyan, mint egy adott állapotban lenni, amit csak úgy lehet megismerni, ha első személyben tapasztaljuk meg, hogy az adott állapotban vagyunk.

Ami vitatott, az pontosan az, hogy a kvaliák milyen magyarázó terhet rónak az elme teoretikusaira, ha egyáltalán rónak ilyen terhet.

A fenti gondolat kísérletet eredetileg Frank Jackson ausztrál filozófus írta le - bár kissé differens formában - 1982-es *Epiphenomenal Qualia* című cikkében. Az itt bemutatott, a tisztán fizikai tudatelméletek magyarázó adekvát volta ellen szóló érvet - amely szintén ott található - tudásérvnek nevezzük. Ezt az érvet hatalmas irodalom övezi, amelyet itt nem kívánok összefoglalni. Az érdeklődő olvasót még egyszer a további olvasmányokra vonatkozó javaslatokhoz utalom.

Azonban nyugodt lehetsz, hogy ez biztosan nem az utolsó, amit a qualia témakörében ebben a kötetben látni fogunk. Egyelőre azonban itt az ideje, hogy figyelmünket a funkcionalizmusról szóló beszámolóra fordítsuk.

6. FEJEZET

FUNKCIONALIZMUS

Ebben a fejezetben nagyon kevés magyarázatot kell adni, ezért viszonylag rövid lesz. Ennek két oka van.

Ennek egyik oka az, hogy a funkcionalizmushoz szükséges előkészületek nagy részét már elvégeztük az elme kauzális elméletének kidolgozása során. Ez volt az egyik indíték arra, hogy az ausztrál materializmussal együtt mutassuk be a kauzális elméletet.

A másik ok az, hogy a funkcionalizmus szigorúan véve egy olyan *elméleti keret*, amelyet ki kell egészíteni egy teljes értékű elmeelméletté. A funkcionalista keretrendszer kiteljesedésének egyik módja adja azt az elméletet, amelynek kidolgozása és értékelése e kötet központi témája - a *komputacionalizmust*. Mielőtt azonban teljes mértékben kifejlesztenénk a komputacionalizmust, felfüggesztjük az elme filozófiai elméletekről szóló vitánkat, és kidolgozunk egy szigorú technikai beszámolót arról, hogy pontosan mi is a *számítás*.

Egyelőre megelégszünk azzal, hogy világossá tesszük az elméleti keretrendszer szerkezetét, amelyet később részletesebben kifejleszteni fogunk.

6.1 FUNKCIONÁLIS MEGHATÁROZÁS

Kezdjük - ahogyan az oksági elmélet esetében is tettük - azzal, hogy elgondolkodunk bizonyos fogalmak meghatározó jellemzőin.

Nyelvünkben sok kifejezést a referensük jellemző ~~függő~~ határoz meg. Ennek paradigmaticus példája a "karburátor" kifejezés. A karburátor a belsőégésű motorok (ma már nagyrészt elavult) eszköze, amelynek feladata az üzemanyag és a levegő pontos arányú keverése a maximálisan teljes égés érdekében. A karburátorok készülhetnek fémötvözetből, kerámiából vagy más anyagból. Használhatnak pillangószelepeket, csúszókarokat vagy más mechanizmusokat a be- és kimenet szabályozására. Lehetnek egy vagy több kamrásak. Bármilyen színűek lehetnek. A felsorolt jellemzők egyike sem,

azonban semmilyen hatással nincs arra, hogy valamit megfelelően "karburátornak" neveznek-e vagy sem.

Valami karburátor iff a karburátor funkcióját látja el - iff közvetít az üzemanyag- és levegőbevitel és az égési keverék kimenete között. Bármilyen, ami ezt a funkciót képes ellátni, karburátornak számít.

A funkcionista, ahogyan mostanra már bizonyára kitalálta, azt vallja, hogy a mentális állapotok kifejezései pontosan ilyen kifejezések. Ami egy állapotot mentális állapottá tesz, az a funkcionista szerint nem valami, ami az állapotnak sajátja, hanem inkább az a funkciója, hogy közvetíti a bemenetek, kimenetek és más mentális állapotok közötti kapcsolatokat. A mentális állapotokat *funkcionális* állapotoknak tartják.

A mentális állapotok bármelyik típusának megértése a funkcionista elemzés szerint egyszerűen a funkció megértésének kérdése. A fájdalmat például olyan funkcionális állapotnak tekintik, amelyet az emberekben jellemzően testi trauma okoz, és amely jellemzően szorongást és a fájdalom enyhítésére irányuló gondolkodást, valamint jellemzően a fájdalom enyhítésére irányuló viselkedést vált ki.

Más szóval, a fájdalom olyan funkcionális állapot, amely közvetíti a kapcsolatot a jellegzetes fájdalmat kiváltó bemenetek, a fájdalmat enyhítő okok és a viselkedés között. Bármilyen, ami alkalmas arra, hogy ezt a funkciót szolgálja.

– a kapcsolatok ilyen módon történő közvetítése - *csak* egy fájdalmas állapot.

Kétségtelenül arra gondolsz, hogy ez nagyon is úgy hangzik, mint az oksági elmélet. Ha így van, akkor minden bizonnyal igaza van - az oksági elmélet valójában a funkcionálizmus egy korai formája.

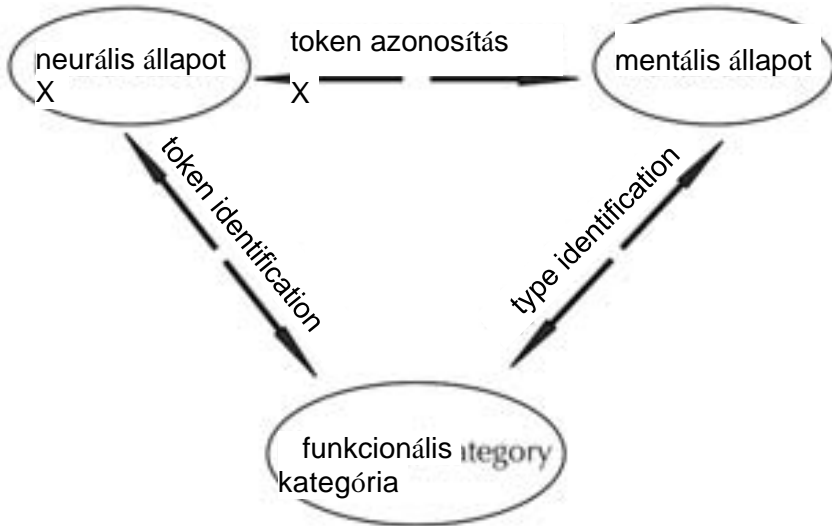
Arra is gondolhat, hogy ez eléggé emlékeztet a viselkedési turizmusra. Ismét igaza van. A funkcionálizmus bizonyos értelemben az új behaviorizmus. A funkcionista beszámoló megőrzi az inger, a mentalitás és a viselkedés közötti fontos kapcsolatot. A döntő különbség azonban az, hogy a funkcionista *nem* eliminativista a mentális állapotok tekintetében. Valójában a funkcionista azt vallja, hogy a mentális állapotok bármely adekvát leírása tartalmaz egy *más mentális állapotokra* való, nem elemezhető utalást.

A funkcionista szerint a mentális állapotok jellemző funkciója az, hogy közvetítik a bemenetek, kimenetek és *más mentális állapotok* közötti kapcsolatokat.

6.2 A FEKETE DOBOZ ELMÉLET

Annak érdekében, hogy az elméleti keretrendszer szerkezetét teljes

mértékben átlássuk, hasznos a leírás három szintjét ábrázolni - és



6.1. ábra Típus/token azonosság.

a közöttük fennálló azonosságokat - diagramszerűen (lásd a 6.1. ábrát).

Amint láthatjuk, a funkcionalista keretrendszerben mind típus-, mind token-azonosítások szerepelnek. Egy adott állapotot - példa kedvéért neurális állapotot használtam - egy adott mentális állapottal azonosítunk. Az ausztrál materializmussal ellentétben azonban ez nem típus-azonosítás. A neurális állapotot a mentális állapottal azáltal azonosítják, hogy egy bizonyos funkcionális szereppel azonosítják, amely maga is típus-azonosítható a mentális állapottal.

Más szóval, a mentális állapotok típusai *éppen* a funkcionális szerepek típusai. A mentális állapotok tehát típusazonosak a funkcionális kategóriákkal. Bármely állapot, amely alkalmas az adott funkció betöltésére, *éppen* mentális állapot, de *nem* az állapot bármely belső tulajdonsága alapján. Ezért az azonosság csak jelképes azonosság. A 6.1. ábrán látható neurális állapot helyén bármi állhat, ha a megfelelő funkciót végzi. Még egy másik módja annak, hogy ezt úgy fejezzük ki, hogy a 6.1. ábrán látható neurális állapot ebben az esetben *történetesen* azonos a mentális állapottal - *történetesen* az a dolog, ami a mentális állapot -, de nem azáltal, hogy a neurális állapotnak valamilyen meghatározott *típusa*.

Bármilyen típusú neurális állapot (valójában bármilyen típusú állapot egyáltalán) *lehet történetesen* azonos a mentális állapottal, ha *történetesen* a szükséges funkciót látja el.

Az előző fejezet végén röviden tárgyalt jelképes fizikalizmussal ellentétben a funkcionalizmus nem módszertanilag üres, a

az empirikus vizsgálat tekintetében. Éppen ellenkezőleg. A mentális állapotok és a funkcionális kategóriák közötti típus-azonosítás révén pontosan tudjuk, hogyan kell empirikusan vizsgálnunk a mentalitást. A pszichológiai vizsgálat a funkcionista elképzelés szerint a mentális állapotok egyes típusainak jellemző funkcióinak meghatározását és vizsgálatát jelenti.

Ez a funkcionista keretrendszer egyik egyértelmű előnye - pontosan úgy irányítja a pszichológiai kutatást, ahogyan azt egy empirikusan adekvát elmeelmélettől elvárjuk. A mentalitás jobb megértése azt jelenti, hogy a mentális állapotok egyes típusai által a jellemző bemenetekhez, kimenetekhez és más mentális állapotokhoz való viszonyuk közvetítésének számbavételét kell kidolgoznunk.

A funkcionizmus további előnye abban rejlik, hogy megőrzi azokat az intuíciónkat, amelyek az elődelméleteket megalapozták. Megmarad a behaviorista intuíció, miszerint a mentalitás döntően az inger és a viselkedés közötti kapcsolatokat foglalja magában, és az ausztrál materialista intuíció, miszerint a mentális életet az idegi aktivitás szempontjából kell megmagyarázni.

Fontos azonban tisztában lenni azzal, hogy bár az ember szabadon azonosíthatja az idegi állapotokat olyan dolgokkal, amelyek a mentális állapotok szerepét töltik be, nem *kötelezi magát* erre pusztán a funkcionizmus iránti elkötelezettség miatt.

A funkcionizmus fontos jellemzője, hogy *szubsztrátfüggetlen*, tehát ontológiailag semleges. Bármilyen - beleértve a nem fizikai szubsztancia állapotát is - lehet mentális állapot a funkcionista szemlélet szerint, feltéve, hogy ellátja a szükséges funkciót. Ez a szubsztrátfüggetlenség pontosan az, ami lehetővé teszi a funkcionista számára, hogy a mentális állapotok többszörös megvalósíthatóságát elfogadják anélkül, hogy módszertani ürességbe esnének.

Ennek a szubsztrátfüggetlenségnek a következménye az ausztrál materializmusban rejlő fajsovinizmus elkerülése. Ami a funkcionista illeti, nyitott kérdés, hogy a nem emberi entitásoknak - biológiai vagy egyéb - vannak-e mentális állapotai vagy sem. A funkcionizmus lehetővé teszi, hogy kutyáknak, macskáknak, marslakóknak és - ami döntő fontosságú - ember alkotta tárgyakkal is legyen elméjük. A funkcionista keretrendszer lehetővé teszi a mesterséges intelligencia lehetőségét.

A funkcionizmus a szubsztrátfüggetlenség mellett a mechanizmustól is független. Semmit sem mond arról a tényleges mechanizmusról, amellyel a mentális állapotok a bemenetek, kimenetek és más mentális állapotok közötti kapcsolatok közvetítésében betöltik funkciójukat. Emiatt a funkcionizmust gyakran nevezik a mentalitás "fekete doboz" elméletének.

A funkcionalista megközelítés szerint a mentális állapotok olyanok, mint a fekete dobozok. Nem tudjuk, hogy miből állnak, és azt sem, hogy mi zajlik bennük. Miközben ez az általunk leírt elméleti előnyökkel jár, van egy olyan értelemben is, hogy a mentális állapotok funkcionalista beszámolója hiányt hagy maga után. Az ember többet szeretne tudni, különösen azoknak a mechanizmusoknak a részleteiről, amelyek elősegítik a mentális állapotokra jellemzőnek tartott közvetítést. Ezért hivatkozom a funkcionalizmusra mint *elméleti keretre*. A szóban forgó mechanizmusok specifikálásának különböző módjai különböző teljes értékű funkcionalista elméleteket eredményeznek.

A *számítás* az első számú jelölt egy olyan mechanizmusra, amely alkalmas az ilyen közvetítés pontos elvégzésére. Ha a funkcionalizmust a közvetítő mechanizmus számítással történő bemutatásával egészítjük ki, akkor megkapjuk azt az elméletet, amely központi kérdésünk - a *komputacionalizmust*.

Ahhoz, hogy ezt felelősségteljesen és pontosan megtehessük, szigorú formális beszámolóra van szükségünk arról, hogy mi is a *számítás*. Ez lesz a következő három fejezet célja.

Mielőtt azonban rátérnénk erre a formális anyagra, röviden vizsgáljunk meg néhány szokásos filozófiai ellenvetést a tág funkcionalista keretrendszerrel szemben.

6.3 QUALIA ELLENVETÉSEK

A két kifogás, amelyet itt felhozunk, nem a funkcionalizmus egy bizonyos fajtáját célozza, hanem inkább a funkcionalista keretrendszer középpontjában álló állítást. Ezek ellenvetések azzal az állítással szemben, hogy a mentális állapotokról a funkciójukon túl semmi más fontosat nem kell tudni, és hogy egy ilyen funkció betöltése *elengedhetetlen* ahhoz, hogy valami mentális állapot legyen.

Mindkét ellenvetés arra törekszik, hogy kiemelje a *kvaliák* fontosságát a mentális életben, és célja, hogy magyarázó terhet rójon az elme teoretikusaira a kvaliák magyarázata érdekében. Vegyük először a *fordított spektrumú* ellenvetést.

Amikor bizonyos fényvisszaverődési tulajdonságokkal rendelkező tárgyak jelenlétében, bizonyos fényviszonyok mellett olyan perceptuális élményem van, hogy ha valaki megkérdezi tőlem, milyen színűnek látom a szóban forgó tárgyat, azt válaszolom, hogy "kék". A Csendes-óceánt erős napfényben figyelve szerzett tapasztalatom egy para- digmikus példa arra, hogy mire utalok a "kék" színekifejezéssel.

Amikor Sue ugyanolyan fényviszonyok között, ugyanolyan tárgyak jelenlétében van, és megkérdem tőle, hogy milyen színt

50

érzékel, szintén azt válaszolja, hogy "kék". Amikor én látok egy tárgyat, amit sárgának érzékelek.

és azt mondja, hogy "ez sárgának tűnik", Sue egyetért. Így van ez az összes többi színkifejezésünkkel is. Más szóval, mindig és mindenhol egyetértünk a színkifejezéseink kiterjesztésében.

A probléma az, hogy nincs közvetlen hozzáférésem Sue érzékelési tapasztalataihoz, csak a tapasztalatairól szóló beszámolóihoz. Mivel mindig és mindenhol ugyanazokra a dolgokra mutatunk, amikor színkifejezéseket mondunk, *feltételezem*, hogy amikor Sue-nak olyan érzékelési élménye van, amelyről "kékként" számol be, akkor neki is *olyan* érzékelési élménye van, *mint* nekem, amikor a kékséget tapasztalom. Amennyire én tudom, azonban az is lehet, hogy amikor Sue azt érzékeli, amit "kék"-ként jelent, akkor *valójában* egy olyan érzékelési élményben van része, hogy ha *nekem* lenne ilyen élményem, akkor azt "sárgának" jelenteném. Valójában, amennyire én tudom, ez lehet a helyzet az összes színkifejezésünk esetében - a színspektrumunk lehet, hogy teljesen fordított egymáshoz képest.

A mentális kifejezések funkcionális elszámolásának megfelelősége ellen szóló érv, amelyet ebből a gondolat kísérletből merítünk, a következő vonalon halad. Az én mentális állapotom, hogy "kéket érzékelek", és Sue mentális állapota, hogy "kéket érzékelek", *funkcionálisan egyenértékű*. A mi megfelelő állapotaink ugyanúgy közvetítik a kékre jellemző kék típusú ingert, a kéket érzékelő viselkedést és más mentális állapotokat. A funkcionista szerint tehát az állapotaink *egyenértékűek*, és nincs semmi, ami megkülönböztetné őket. Végül is, a funkcionista szemlélet szerint a mentális állapotok közötti egyetlen lehetséges különbségtétel a funkcióban való különbségtétel. Teljesen világosnak tűnik azonban, hogy a "kéket érzékelő" állapotainkban van valami egészen más. Sue, az ő fordított színspektrumának köszönhetően, azt érzékeli, amit én "sárgának" neveznék.

Egyáltalán nem világos, hogy mit mondhatnánk erről az érvről, és természetesen nem áll szándékomban itt döntenem róla. A szakirodalom, mint már említettem, nagyon megosztott a qualia fontosságát illetően.

Egyrészt egyértelműnek tűnik, hogy a szóban forgó két mentális állapot között fontos különbség van, amely nem magyarázható pusztán a funkció szempontjából. Az, hogy milyen számomra a kéket érzékelni, egyáltalán nem olyan, mint az, hogy milyen Sue számára a kéket érzékelni. A két élményhez egészen eltérő kvalitások kapcsolódnak.

Másrészt nem nyilvánvaló, hogy mi függ ettől. Végül is, ha mindig és mindenhol egyetértünk a színkifejezéseink kiterjesztését illetően, akkor bizonyára csak ez a fontos. Tényleg számít, hogy én a "kék" élményét "sárga" élménynek nevezném? Különösen, mivel ez

olyasvalami, amit sem mi, sem más nem tudhat soha?

A mentalitás funkcionális leírásának megfelelőségével szembeni második ellenvetés a *hiányzó kvalitások* ellenvetése, vagy - még szórakoztatóbban - a *zombi ellenvetés*.

Ha úgy tetszik, gondoljatok egy lényre, aki minden értelemben és céllal megkülönböztethetetlen töletek és tölem. Nevezzük ezt a lényt Imitációs Embernek. Az Imitált Embernek normális élete van, akárcsak neked és nekem. Vannak kedvencei és ellenszenvei, céljai és ambíciói, hite és vágyai - egyszóval olyan, mint bármelyik másik emberi lény. Ha megkérdezzük az élményeiről, az Imitációs Ember olyan válaszokat fog adni, amelyeket bármelyik másik embertől elvárnánk - elmondja, hogy fájdalmat érez, hogy bizonyos zenehallgatás élvezetes, és hogy a csokoládéfagylalt íze csodálatos.

Veled és velem ellentétben azonban az Imitation Man teljesen nélkülözi a kvalitásokat. Őt nevezhetnénk "zombinak". A fájdalom valójában nem *érez* semmit. Nincs semmi olyan, ami számára olyan lenne, *mintha* zenét hallgatna, vagy fagylaltot kóstolna. Nincs rá mód, hogy ezt valaha is felfedjük, mivel természetesen nincs közvetlen hozzáférésünk az ő tapasztalataihoz.

Ez semmiképpen sem szól a többi emberrel való *funkcionális egyenértékűsége* ellen. A fájdalomállapota még mindig ugyanazt a funkcionális szerepet játssza, mint a fájdalomállapotok mindenki másnál, ahogyan az összes többi mentális állapota is. Egyszerűen csak nem *érezik magukat* semminek.

A funkcionalizmus ellen szóló érvnek nyilvánvalónak kell lennie. Az utánczó ember fájdalomállapota például funkcionálisan egyenértékű a miénkkel - ugyanúgy közvetíti az inger, a viselkedés és más mentális állapotok közötti kapcsolatokat, mint a mi fájdalomállapotunk. Ha az Imitált Ember kezét a kályhára teszi, akkor kiabálni fog, felkiált, hogy "áúú" (vagy más megfelelő káromkodást), megpróbálja eltávolítani a kezét a hőforrástól, kéznyugtató viselkedésbe kezd, sír és nyög, és így tovább. Az utánczás Az ember fájdalom és a mi fájdalmunk tehát a funkcionalista szemlélet szerint *simpliciter* egyenértékű. Mégis egyértelműnek tűnik, hogy az Imitált Ember fájdalomja meglehetősen differens dolog a mi fájdalmunktól - az ő fájdalomja nem *fáj*.

A fordított spektrummal kapcsolatos ellenvetéshez hasonlóan egyáltalán nem világos, hogy mit kellene mondanunk a hiányzó kvalitással kapcsolatos ellenvetésre. Ismétlem, itt nem fogok dönteni róla, hanem csak jelzem, hogy milyen kétféle módon lehetünk hajlamosak válaszolni.

Egyfelől úgy tűnik, hogy a példa igazságtalanul prejudikál a funkcionális magyarázat megfelelőségével szemben. Feltételezi, hogy az emberi mentális életre jellemző különféle funkciók elvégzése nem *ipso facto* rendelkezik qualia-val. Valójában csábító a funkcionalista ellenzöt azzal vádolni, hogy a kérdés felvetése a

54

funkcionalista ellen szól.

Azonban a funkcionalista beszámolóban eddig semmi olyan nincs, ami a kvalitások jelenlétét a mentalitásra jellemző különböző funkciók elvégzésének egyenes következményeként tartaná fenn.

A funkcionalista csak azt állítja, hogy a mentális állapotok funkcionális beszámolója *megfelelő* beszámoló, így a hiányzó qualia ellenzője nem *egészen így teszi* fel a kérdést. Azt azonban difficult értékelni, hogy az Imitált ember fájdalomállapota funkcionálisan hogyan lehetne egyenértékű a miénkkel anélkül, hogy a fájdalom fájdalmas lenne. Végül is éppen a fájdalomunk fájdalmas volta motiválja a fájdalomcsillapító viselkedésünket, nemde? Még mindig van olyan értelemben, hogy a hiányzó kvalitással kapcsolatos kifogás valahogy a funkcionalista ellen szól.

Másfelől a hiányzó kvalitások ellenvetése úgy tűnik, hogy *pontosan* a mentális állapotok funkcionális leírásának elégtelenségét mutatja. A mentalitás funkcionalista beszámolója nem képes megragadni mentális állapotaink lényegesnek tűnő aspektusait - azok szubjektív minőségi aspektusait. *Egyszerűen van* valami, ami olyan, mint kéket látni, vagy fájdalmat érezni, vagy csokoládéfagylaltot kóstolni, és azt gondolhatnánk, hogy minden olyan elmélet, amely nem ad számot az ilyen kvalitásokról, magyarázatilag elégtelen.

Ezek a diffikult filozófiai kérdéseken való elmélkedés után itt az ideje, hogy figyelmünket a számítás formális elszámolása felé fordítsuk.

FORMÁLIS RENDSZEREK

Az előző fejezetekben azt a kérdést vizsgáltuk, hogy mi *lehet az* elme, és felvázoltuk az erre a kérdésre adott lehetséges válaszok terét. Ennek során láttuk az elme filozófiai elméleteinek fejlődését, és megvizsgáltuk az egyes elméletekhez tartozó érveket és ellenvetéseket.

A következő három fejezetben a *számítógépek* pontos formális leírása felé fogunk haladni.

Ellentétben azzal a kérdéssel, hogy mi lehet az elme - amely megérett a te- orizációra -, van valami, *ami* számítógépnek lenni, és ennek a valaminek a meghatározása tisztán leíró jellegű feladat, amely magában foglalja az elméleti informatikában való elmélyülést és néhány alapanyag kiemelését.

Ezekben a fejezetekben nem feltételezem, hogy bármit is értenék az informatikából, a matematikából vagy bármely formális tudományágból. Ha idegenkedik a szimbólumoktól, akkor ne féljen. A bevezetés szándékosan lassú és szelíd, és számos gyakorlat segíti a megértést.

Ebben a fejezetben a *formális rendszerek* meghatározásával kezdjük, és néhány (egyszerű) formális játékszisztemmel fogunk játszani, hogy megismerjük a szimbólumkezelés alapérzetét. A következő fejezetben a formális rendszerek egy speciális fajtáját fogjuk megvizsgálni: a *regisztergépet*. A regisztergép fogalmát - és a kifejtésében részt vevő kapcsolódó fogalmakat (mint például a *program*) - a *kiszámíthatóság* pontos jellemzésére fogjuk használni.

A kiszámíthatóság e szigorú definíciójának birtokában aztán tekintélyesen (és helyesen) beszélhetünk a *számítástechnikáról*, a *számítógépekről* és a *számításról*. Játszani fogunk néhány játékszisztem-regiszter-gépes programmal, hogy megérezzük a számítás szintaktikai természetét. Megnézünk néhány diffiktabb és bonyolultabb regisztergépes programot is (azoknak, akik fogékonyak az ilyesmire és szeretik a kihívásokat). Ezek a diffiktabb kihívást jelentő

a gyakorlatok előítélet nélkül kihagyhatók azok számára, akiknek nincs ízlésük hozzájuk.

Végül a 9. fejezetben befejezzük a számításelméletről szóló áttekintésünket azzal, hogy megnézzük, hogyan használhatjuk a *Gödel-kódolás* nagyon okos módszerét egy *univerzális gép* - egy olyan gép, amely *bármilyen* kiszámítható függvényt képes kiszámítani - definiálására. Azt is meg fogjuk vitatni, hogy mi az, ami kiszámítható - mi tartozik a kiszámítható határain belülre, és mi kívülre.

A számításelmélet alapos ismereteivel felvértezve pontos formális definíciók és néhány finom megkülönböztetés áll majd rendelkezésünkre. Ezeket alkalmazva képesek leszünk helyesen és felelősségteljesen jellemezni azt az elméletet, amely a központi témánk. Ez lesz az első célunk a 10. fejezetben.

7.1 HATÉKONYSÁG

Nagy valószínűséggel e könyv minden olvasója játszott már életének egy szakaszában legalább egy játékot a következők közül: sakk, dáma, backgammon, go, kínai dáma vagy - legalábbis - tic-tac-toe (más néven téglá és kereszt). Ha érted, hogyan játszanak legalább egy ilyen játékot (a legtöbben értjük a tic-tac-toe-t), akkor függetlenül attól, hogy mennyire jól vagy rosszul játszol velük, *már érted* a formális rendszerek alapelveit. A formális rendszerek vizsgálatát azzal kezdjük, hogy egyszerűen explicitté tesszük azt, amit implicit módon már megértettünk.

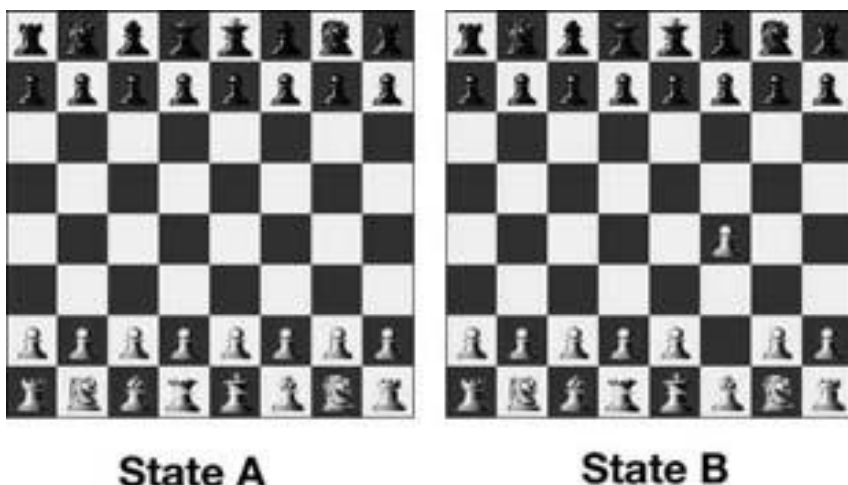
A sakk jól példázza a formális rendszerek fontos jellemzőit, ezért ebben a fejezetben végig hivatkozni fogok rá. Ne aggódjon, ha nem érti különösebben a sakk szabályait vagy a mögötte álló stratégiát - semmi, amit mondani fogok, nem függ ettől a megértéstől.

A formális rendszerek jellemzőinek felvázolásához tekintsük a 7.1. ábrán látható sakktábla-konfigurációkat.

Azonnal látható, hogy a két lap differens *állapotban* van. Ráadásul mindannyian egyetérthetünk abban, hogy a két ábrázolt állapot hogyan különbözik: az egyik fehér bábu két mezőt lépett a fekete bábu felé. Ennél azonban pontosabbak is lehetünk. Ha a vízszinteset 'a'-tól 'h'-ig balról jobbra, a függőlegest pedig '1'-tól '8'-ig alulról felfelé jelöljük, akkor azt mondhatjuk:

[1] Az A állapotban az f2-es mezőben lévő bábu a B állapotban az f4-es mezőben van.

Ha ismerjük a sakkot, akkor az [1] értelmezési rétegeket is hozzáadhatunk. Az értelmezés első szintjén azt mondhatjuk, hogy egy *gyalog*, amelyik a



7.1 ábra Két állapot.

az A állapot f2 négyzete a B állapot f4 négyzetében van. Az interpretáció következő szintjén azt mondhatjuk, hogy az A állapotból a B állapotba való átmenet egy *érvényes* sakk lépés - a játékszabályoknak megfelelően végrehajtott lépés. Azt is mondhatjuk, hogy az A állapot a sakkjátzsma kezdeti konfigurációját, azaz *kezdeti* állapotát jelenti. Az értelmezés egy másik szintjén azt mondhatjuk, hogy az ábrázolt lépés *érdekes* vagy *unalmas* lépés. Mindezek az értelmezések azonban egyelőre nem érdekelnek minket.

Mi csak az [1]-re akarunk koncentrálni, és arra használjuk, hogy kiemeljük a sakk néhány alapvető jellemzőjét, amelyek nélkül a játékot nem lehetne játszani - olyan nyilvánvaló tulajdonságok, amelyeken valószínűleg soha nem volt okod elgondolkodni.

Az [1]-ben az érdekel minket, hogy egy A állapotú és az általunk leírt módon felcímkézett sakktábla esetén az [1] tartalmazza az összes szükséges információt a B állapot újraalkotásához - még akkor is, ha *egyáltalán nem* tudunk *semmit* a sakkról. Valójában még csak fel sem kell ismernünk az állapotokat egy sakktábla konfigurációjaként ahhoz, hogy az [1]-ben szereplő információt az A állapotról alkalmazzuk, és *létrehozzuk* a B állapotot.

Fogalmazzuk át az [1]-et egy feladat vagy eljárás szempontjából a következőképpen:

[2] Vegyük az f2 mezőn lévő figurát, és vigyük az f4 mezőre.

Ha egy A állapotba konfigurált, felcímkézett sakktáblát kapunk, és azt mondjuk [2], könnyen teljesíthetjük a feladatot anélkül, hogy megértenénk a feladat *jelentőségét*, anélkül, hogy értelmeznénk annak *jelentését*. Nyilvánvaló, hogy a feladatot *leíró* szavak jelentését természetes nyelven kell értelmeznünk, de maga a feladat

elvégezhető

pusztán mechanikusan, a jelentőségének értékelése nélkül. Az ilyen feladatot *effektív eljárásnak* fogjuk nevezni.

A [2] A állapotra való alkalmazása a B állapot létrehozásához a [2] *effektivitása azon a tényen alapul, hogy a sakkjátszma állapotai effektív módon megkülönböztethetők.* Az effektivitás fogalmával a következő fejezetekben fogunk némi munkát végezni, ezért tárgyaljuk a sakk azon nyilvánvaló tulajdonságait, amelyeket említettem, és amelyek alapján a sakkjáték állapotai effektíven megkülönböztethetők.

Először is, nincs kétértelműség arról, hogy hol kezdődik vagy végződik a bábú, sem arról, hogy hol kezdődik vagy végződik a négyzet. A darabok különállóan körülhatárolt térbeli objektumok, a négyzetek pedig világosan körülhatárolt határokkal rendelkeznek. Másodszor, soha nincs kétség afelől, hogy egy bábú "benne van-e" egy négyzetben vagy sem. Ezek a feltételek szükségesek ahhoz, hogy létezzenek olyan *formális szabályok*, amelyek szabályozzák a bábuk mozgásának törvényes módjait - ami a sakk és minden hasonló játék jellemzője.

Ha azt mondjuk, hogy a sakk állapotai *effektíven megkülönböztethetők*, akkor azt mondjuk, hogy létezik egy *effektív eljárás*, amely eldönti a kérdést. Mostanra már valamennyire érzékelnie kellene, hogy mi lehet egy effektív eljárás - a [2] egy példa erre. Próbáljunk meg egy munkadefiníciót kidolgozni. Nevezzünk egy eljárást effektívnek, *iff* ha pusztán egy meghatározott lépéssorozat - egy utasításlista, egy recept - követésével érhető el, anélkül, hogy a feladat jelentőségét vagy értelmét megértenénk. Ezt az utasításlistát nevezhetjük *algoritmusként*. Egy másik módja annak, hogy egy eljárás effektivitására utaljunk, az, hogy azt mondjuk, hogy az eljárás *algoritmikus* - hogy van egy algoritmus az eljárás végrehajtására.

Az "algoritmikus" kifejezés a köznyelvben használatos - hallhatta már, hogy valaki egy feladatot vagy feladatot tisztán algoritmikusként nevez, hogy leírja, hogy egyszerűen csak a mozdulatok elvégzéséről van szó, az előírt lépések megtételéről - egy hétköznapi feladatról. Más szóval, egy effektív eljárás.

Elég egyértelmű, hogy a sakkállapotok megkülönböztetése algoritmikus.

Adjuk meg az algoritmust:

1. Kezdjük az a1 négyzettel.
2. Ismételve meg a 3. lépést hatvannégyszer, hacsak nem kap utasítást a leállításra. Ha azt az utasítást kapjuk, hogy lépünk a következő négyzetre, lépünk a jobb oldali szomszédos négyzetre, ha van ilyen, különben lépünk a közvetlenül felette lévő sor legelső bal oldali négyzetére.

3. Ha az A állapotban lévő négyzet üres, akkor:
 - ha a B állapot megfelelő négyzete üres, lépünk tovább a következő négyzetre, ellenkező esetben álljunk meg, és mondjuk ki, hogy "a két állapot different".

Ha az A állapotban lévő négyzet foglalt, akkor:

- ha a B állapot megfelelő négyzetét ugyanolyan alakú figura foglalja el, lépjen tovább a következő négyzetre, ellenkező esetben álljon meg, és mondja ki, hogy "a két állapot different".

4. Mondja ki, hogy "a két állapot formailag egyenértékű".

Bár ennek az algoritmusnak a meghatározása egy olyan feladat meglehetősen fárasztó leírását jelenti, amelynek végrehajtása számunkra egyszerű és nyilvánvaló, mégis azt a pontot hangsúlyozza, hogy a feladat elvégzéséhez nincs szükségem semmilyen ismeretanyagra. Nem kell tudnom, hogy sakkállapotokat hasonlítok össze. Nem kell tudnom, hogy mi a sakk. Még azt sem kell tudnom, hogy mi a játék. Csak követnem kell a *formailag meghatározott* utasításokat.

Azt gondolhatod, hogy ez az algoritmus nem tükrözi jól, hogyan képzeled el magad a feladat elvégzésében. Ha előre megküldjük az A és a B állapotot, és megkérjük, hogy hasonlítsuk össze őket, azonnal nyilvánvalóvá válik számunkra, hogy hogyan különböznek (anélkül, hogy az összehasonlítást az a1-es négyzetnél kezdenénk, továbblépnénk a2-re stb.), még akkor is, ha csak egy darab elhelyezkedése tekintetében különböznek. Ez részben azért van így, mert az A állapot egy szabályos minta egy sufficiens kis skálán. Ha a bábukat véletlenszerűbben osztanánk el a táblán, kétségtelenül jobban követnénk a fenti algoritmust. Ha a tábla mindkét oldala mentén tízszer hosszabb lenne - 6400 négyzet területű -, és tízszer annyi bábu lenne, akkor a két állapot helyes összehasonlítására csak a fenti algoritmus követése lenne az egyetlen reményünk.

Mindenesetre itt csupán arról van szó, hogy a feladat egy algoritmus követésével megvalósítható, és ezért effektív.

Az effektivitás utolsó szempontja, amelyet értékelnünk kell, az effektív eljárások *végességre* vonatkozó *lovasa*. Egy eljárás csak akkor számít effektívnek, ha *véges idő* alatt végrehajtható.

A molekulák megszámlálása ebben a könyvben egy effektív eljárás. Jelenleg mindannyian egy szilárd testben vannak elrendezve, és nagyjából mozdulatlanul ülnek, így elvileg felállíthatnánk egy nagyon finom háromdimenziós rácsot referenciapontként, és egyszerűen sorra megszámlálhatnánk az egyes rácskockákon lévő molekulákat. Ez nagyon-nagyon sokáig tartana, de *véges* ideig.

A természetes számok megszámlálása viszont *nem* egy effektív eljárás. Bár a folyamat mechanikus - *nevezzük meg a nullát; nevezzük meg az utoljára megnevezett szám utódját; ismételjük meg az előző lépést, amíg tovább nem jutunk* -, véges idő alatt nem fejezhető be. Minden természetes számnak van utódja - végtelen sok van

belőlük. Tehát ennek a folyamatnak a végrehajtása nem effektív.

Most, hogy már jól megértettük az effektivitást, itt az ideje, hogy a fogalmat a formális rendszerek jellemzésében alkalmazzuk.

7.2 ÁLLAMOK ÉS SZABÁLYOK

A formális rendszerek két gyűjteményből állnak: *állapotok* gyűjteményéből és *szabályok* gyűjteményéből. Egy adott formális rendszer specifikációja az állapotok és a szabályok specifikációjából áll.

Bármely általunk választott entitáson keresztül definiálhatunk állapotokat, feltéve, hogy a két adott állapot közötti különbségtétel effektív.

Hogy néhány példát említsünk, az állapotokat definiálhatjuk egy játéktábla konfigurációjaként, kapcsolók véges tömbjének konfigurációjaként, emberek eloszlásaként a színházi ülések véges tömbjében, vagy levelek eloszlásaként a galambdobok véges tömbjében.

Csak véges számú *entitáson* lehet állapotokat definiálni, különben két állapot nem lesz effektíven megkülönböztethető. A gyűjteményben azonban végtelen sok *állapot is* lehet - sok formális rendszer, amellyel találkozni fogunk, végtelen sok olyan állapottal rendelkezik, amelyeket csak néhány entitáson keresztül definiálnak *rekurzívan*. Mindjárt meglátjuk, hogyan valósul meg ez a trükk.

A fenti példák mindegyike fizikai tárgyakat használ *tokenekként* - darabokat, kapcsolókat, embereket, leveleket -, és az állapotokat e tokenek diszkrét elrendezése különbözteti meg. Ugyanilyen könnyen használhatunk azonban papíron lévő szimbólumokat is tokenekként, és definiálhatjuk az állapotokat e szimbólumok sorozataival (feltéve persze, hogy a sorozatok effektíven megkülönböztethetők). Az összes formális rendszer, amellyel a számításelméleti áttekintésünk során játszani fogunk, éppen ilyen *szimbólumrendszer* lesz.

Miután meghatároztuk az állapotok gyűjteményét, meg kell adnunk a *szabályok* gyűjteményét. A szabályok az állapotokkal operálnak, hogy más állapotokat *hozzanak létre*. A szabályokat, akár csak az állapotokat, csak az effektívitas szempontjai korlátozzák. Állapotokat bármi felett definiálhatunk, feltéve, hogy bármelyik adott kettő effektív módon megkülönböztethető. Hasonlóképpen, a szabályok is lehetnek bármi, amit csak akarunk, feltéve, hogy két megkötésnek megfelelnek.

Először is, annak meghatározásának, hogy egy adott szabály egy adott állapotra vonatkozik-e, effektívnek kell lennie. Jellemzően nem minden szabály vonatkozik minden állapotra. Másodszor, ha egy szabály vonatkozik egy állapotra, akkor effektív módon a

lehetséges kimeneti állapotok véges halmazát kell szolgáltatnia.

A szabályok tehát állapotokat vesznek, effektív módon módosítják azokat, és különböző állapotokat adnak vissza.

Most, hogy már tudjuk, hogyan kell egy formális rendszert specifikálni, és milyen korlátozások vonatkoznak az állapotokra és szabályokra, példázzuk ezt az első játék formális rendszerünkkel.

7.3 MŰSZAKI ADATOK

A formális rendszerek tulajdonságait és műveleteit szeretnénk megvitatni, ezért definiáljunk egy *szimbólumrendszert*, amivel játszhatunk. Nevezzük [STR]-nek - ez egy *karakterláncrendszer* lesz.

A rendszer állapotai a ■ és a . Néhány példa: ■■ a [STR] egyik állapota, ■■■■ a [STR] egyik állapota és a [STR] egyik állapota.

Elég világosnak kell lennie, hogy bármely két állapot effektíve megkülönböztethető. Az már talán nem ennyire világos, hogy miközben csak kétféle szimbólum van, amelyek felett állapotokat definiálunk, és miközben ezekből csak véges sok *token* lehet bármelyik állapotban, mégis *végtelenül sok* állapot létezik. Ez talán világosabb lesz egy pillanat múlva, amikor az állapotok *formális* specifikációját adjuk meg rekurzív definíció segítségével; először azonban még néhány fogalomhoz kell hozzásegítenünk magunkat.

Először is szükségünk van a *kezdeti állapot* fogalmára. A formális rendszerek specifikációi gyakran tartalmaznak egy kezdeti állapotot, vagy a rendszer kezdeti konfigurációját. Minden társasjátéknak van kezdeti állapota - a 7.1. ábrán az A állapot a sakk kezdeti állapotát jelenti.

Másodszor, van egy különleges és fontos állapota ennek a rendszernek (és általában a karakterláncrendszereknek), ez pedig az *üres karakterlánc*. Az üres karakterlánc pontosan ez - egy olyan karakterlánc, amely egyáltalán nem tartalmaz szimbólumokat. Ezt \emptyset -nak fogjuk írni. Fontos megjegyezni, hogy az üres karakterlánc *egy karakterlánc*, és - a kikötés szerint - az [STR] egyik állapota. Az üres karakterlánc szükségessége egy pillanat múlva nyilvánvalóvá válik.

Szükségünk lesz *string változókra* is, amelyeket a következő módon írunk le

<I vagy 'f. A string változók stringeket jelölnek - bármilyen tetszőleges stringet (beleértve az üres stringet is). A sztringváltozókat és az üres sztringet azért kell használnunk, hogy formálisan megadhassuk a sufficient általános szabályokat. Ez is nyilvánvalóvá fog válni egy pillanat múlva.

Végül, meg kell jelenítenünk a *karakterláncok összekapcsolását*. A sztring-összekötés egyszerűen két sztring összekapcsolását jelenti - azaz egymás után írjuk őket. Két karakterlánc <I és 'f összekapcsolását <I'f-ként fogjuk írni - azaz tipográfiai összekapcsolásukat. Ha tehát <I a ■■ és 'f a □ karakterláncot jelöli, akkor

$\langle I'f \text{ lesz } \blacksquare\blacksquare\blacksquare\blacksquare \text{ és } 'f\langle I \text{ lesz } \square\square\square\square.$

Vegyük észre, hogy $\emptyset'f\langle I = 'f\emptyset\langle I = 'f\langle I\emptyset = 'f\langle I.$ Más szóval, bármely karakterlánc megegyezik az üres karakterlánccal való összekapcsolásával, függetlenül attól, hogy a

ahová tetted. Egyszerűbben fogalmazva, ha semmit sem raksz rá, mindig marad az, amivel kezdted.

Mielőtt megadnánk a [STR] teljes formális specifikációját, először csak informálisan írjuk le, hogyan fognak működni a rendszer szabályai.

[STR] csak két szabály lesz. Az első szabály azt fogja mondani, hogy vehetünk bármilyen állapotot, amely két dobozzal kezdődik és egy gyémánttal végződik (és a kettő között bármi lehet - beleértve a semmit is -, és kiadhatunk egy olyan állapotot, amely egy gyémánttal kezdődik, és amelyet bármi követ, ami a két doboz és a gyémánt közé került a bemeneti állapotban (ami lehet semmi).

A második szabály azt fogja mondani, hogy vehetünk bármilyen állapotot, amelyben valahol van egy gyémánt (és a gyémánt előtt és után bármi van, amit akarunk - beleértve a semmit is -, és kiadhatunk egy olyan állapotot, amely két dobozzal kezdődik, egy gyémánttal végződik, és a kettő között bármi van, ami a gyémánt előtt volt a bemeneti állapotban.

A szabályok ilyen informális magyarázata meglehetősen fáradságos. A formális specifikáció azonban elég tömör és rendezett, ahogy a következőkből kiderül.

[STR]

[S1]øegy állapot

[S2]Ha'f egy állapot, akkor 'f■ és 'fᅇ ισ αζ.

[S3] kezdetiállapot: ■ ■

[R1]■■■

'fᅇ 'f

[R2]'f<I

■■■'fᅇ

Most gondosan értelmezzük ezt a formális specifikációt, és győződjünk meg arról, hogy az informálisan leírt rendszert megragadja.

Az első dolog, amit a formális specifikációval kapcsolatban meg kell jegyeznünk, hogy míg a rendszer informális leírása sok szószátyár para-gráfot igényelt, addig a formális specifikáció öt rövid sorban van megadva. A formalizálásban nagyfokú gazdaságosságot lehet elérni a kifejezésmódban.

Első pillantásra talán nem egyértelmű, hogy az [S1] és [S2] a rendszer *összes és csakis egyetlen* állapotát rögzíti - ez a *rekurzív definíció* egyik példája. Az [S1] az *alapklauzula* - egyszerűen csak azt írja elő, hogy az üres karakterlánc állapotnak számít. Minden munkát az [S2] - a *rekurzív klauzula* - végez, amely azt mondja ki, hogy ha veszünk egy tetszőleges állapotot, és hozzáadunk egy ■ vagy egy a-t, akkor az eredmény egy állapot lesz. Egy kis gondolkodással meg kell mutatni, hogy az [S1] alapján az ■ és ■ szimbólumok bármely véges karakterláncra előállítható az [S2]

ismételt alkalmazásával. Az [S3] csupán a kezdeti állapotot írja elő.

A figyelmes olvasó elgondolkodhat azon, hogy az [S1] és az [S2] *csak a* rendszer állapotait ragadja meg - az aggodalom oka, hogy úgy tűnik, nem zárják ki kifejezetten a végtelen hosszúságú karakterláncokat. Nem kell azonban aggódnunk - ha bármilyen véges hosszúságú karakterláncot veszünk, és hozzáadunk egy szimbólumot, az eredmény mindig véges hosszúságú karakterlánc lesz.

Az [R1] és [R2] kötetlen olvasatát már adtuk. Addig dolgozzanak oda-vissza e két bekezdés és a fenti formális specifikációjuk között, amíg meg nem győződnek arról, hogy az [R1] és az [R2] valóban megragadja e bekezdések tartalmát. Az egyes szabályok nyílnak bal oldala (LHS) a bemeneti állapotok *formáját*, a jobb oldala (RHS) pedig a kimeneti állapotok *formáját* írja le.

Most már világosnak kell lennie, hogy miért akarunk string változókat használni. A karakterláncváltozók használatával olyan karakterláncok *osztályára* hivatkozhatunk, amelyeknek ugyanaz a *formájuk* (pl. bármelyik karakterlánc, amelyik . . . -val kezdődik és . . . -val végződik). A stringváltozók *mindig* ugyanarra a stringre hivatkoznak a szabály LHS és RHS részében - pontosan ez a használatuk értelme. Bármit is veszünk "f"-nek egy szabály bemeneti oldalán, ugyanannak kell lennie a szabály kimeneti oldalán is. A szabályok *közötti* stringváltozókra azonban nincsenek korlátozások: az, hogy egy szabály alkalmazásakor az 'f' egy stringnek tekinthető, nem befolyásolja, hogy mi lehet az 'f' a szabályok további alkalmazásakor, vagy akár ugyanannak a szabálynak a további alkalmazásakor. Ha ez még nem világos, akkor egy pillanat múlva azzá válik, amikor a rendszer működését vizsgáljuk.

Talán még mindig nem teljesen világos, hogy miért akarunk az *üres karakterláncra* hivatkozni, ezért mutassuk be a használatát.

Az [R1] szabály az ■■ kezdeti állapotra vonatkozik. Ebben az esetben, ami a kezdeti ■■ és a végállapot között van, az egyetlen , így a szabály alkalmazásakor 'f'-nek vesszük , amely esetben az [R1] kimenete . Másképpen fogalmazva: a kezdeti állapot az [R1] alkalmazásához szükséges *alakú*, nevezetesen az ■■'fₓ alakú (ahol ebben az esetben 'f' =). Ha tehát a szabályt alkalmazzuk, akkor a következő kimeneti állapotot kapjuk

◇"f" (ahol "f" =), nevezetesen .

Az [R1] szabály az ■■ állapotra is vonatkozik, mivel az a következő formájú ■■ ■'fₓ (ahol 'f' = ∅). Hasonló érveléssel [R2] az , ■ és □ állapotokra is érvényes lesz. Ezért van szükségünk az üres karakterláncra.

A fenti bekezdés négy szabály alkalmazását említi az államokra: [■■ állapotra és [R2] az , ■ és ■ állapotokra. ◇ ■.

Mi lesz ezeknek az alkalmazásoknak a kimenete az egyes esetekben? Hogyan instanciálódnak a stringváltozók az egyes esetekben?

Emlékezzünk arra, hogy jellemzően nem minden szabály vonatkozik minden állapotra - annak meghatározása, hogy egy adott szabály vonatkozik-e egy adott állapotra vagy sem, effektívnek kell lennie. Nyilvánvalónak kell lennie, hogy sem az [R1], sem az [R2] nem vonatkozik az [STR] minden állapotára. Ugyanilyen nyilvánvalónak kell lennie annak is, hogy annak meghatározása, hogy bármelyik szabály vonatkozik-e egy adott állapotra, effektív.

Ha úgy érzi, hogy nem teljesen követte a szakasz teljes tartalmát, akkor nézze át újra az anyagot, amíg nem érzi magát otthonosan. Az elmúlt néhány oldalon sok új fogalommal találkoztunk, és elkezdtük használni a szimbolikus ábrázolásokat. Erre a megértésre fogunk építeni az elkövetkező oldalakon, ezért a következőkben fontos, hogy először elsajátítsd az eddigi anyagot.

Ha jól érzi magát az eddig tárgyalt anyaggal, és nem okozott nehézséget a 7.1. gyakorlat, akkor itt az ideje, hogy folytassuk, és az [STR] műveleteit használjuk további fogalmak szemléltetésére.

7.4 GENERÁLÁS ÉS SZÁRMAZTATÁS

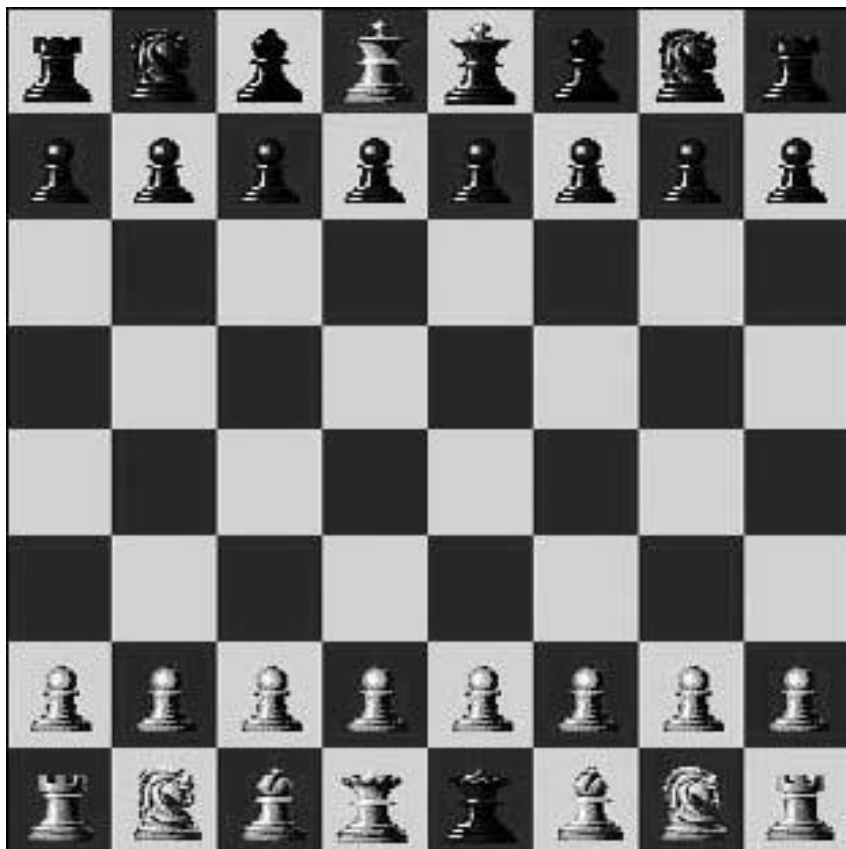
A formális rendszerek műveletei a szabályoknak az állapotokra való egymást követő alkalmazásából állnak. Adott egy kezdeti állapot, és segíthetünk magunknak megkülönböztetni azokat az állapotokat, amelyek a rendszer működése során létrejönnek, és azokat, amelyek, bár megfelelnek a lehetséges állapotok *kritériumainak*, valójában soha nem jönnek létre a rendszer működése során.

Vegyük például a sakkot. A rendszer állapotai harminckét (vagy annál kevesebb), tizenkét (vagy annál kevesebb) típusú - bizonyos megszorításokkal rendelkező - harminckét (vagy annál kevesebb) bábú konfigurációi egy nyolcszor nyolc mezőben. Létezik azonban nagyon sok olyan lehetséges állapot, amely egy sakkjátszma során soha nem fordul elő. Például a 7.2. ábrán látható állapotot a sakk szabályai alapján lehetetlen elérni a kiindulási helyzetből.

Egy állapotot *generált állapotnak* nevezünk, ha az a kezdeti állapotból a szabályok egymást követő alkalmazásával nyerhető. A generált állapotok tehát mindig valamilyen szabály kimenetei lesznek. Következésképpen van egy egyszerű effektív eljárás egy állapot generáltnak való *kizárására*. Ha egy állapot generált, akkor legalább egy szabály kimeneti formájába beleillik, tehát (kontrapozícióval) ha egy adott állapot nem illik bele egyetlen szabály kimeneti formájába sem, akkor *nem* lehet generált állapot.

Sajnos annak meghatározása, hogy egy állapot generálódik-e, nem ilyen egyszerű. Egy szabály kimeneti formájának illesztése nem garantálja, hogy egy állapot generált - ez csupán egy *szükséges feltétel* a generált állapotokra: ennek meghiúsulása garantálja, hogy

nincs generált állapotunk, de teljesülése nem garantálja, hogy van generált állapotunk. Ha ez nem nyilvánvaló, íme egy analóg helyzet. Ha Melbourne-ben vagyok, akkor Ausztráliában vagyok. Tehát az, hogy Ausztráliában vagyok, *szükséges*



7.2 ábra Egy nem generált állapot.

Melbourne-ben való tartózkodás feltétele. Ha tehát nem vagyok Ausztráliában, akkor garantáltan nem leszek Melbourne-ben. Ugyanakkor az, hogy Ausztráliában vagyok, nem garantálja, hogy Melbourne-ben *vagyok* - lehetek Melbourne-n kívül, de Ausztrálián belül is, például Brisbane-ben.

Amikor formális rendszerek iránt érdeklődünk, akkor annak meghatározása érdekel bennünket, hogy bizonyos állapotok *létrejönnek-e* vagy sem. A generált állapot olyan állapot, amely a rendszerben *levezethető*. Ha megmutatjuk, hogy egy rendszerben levezethetünk egy állapotot, az azt jelenti, hogy megadunk egy *levezetést*.

A levezetés a szabályoknak az állapotokra való egymást követő alkalmazásának bemutatása, a rendszer kezdeti állapotától kezdve a minket érdeklő állapotig. Formálisan a levezetés sorok véges sorozata, amelyek közül az első a rendszer kezdeti állapota, a többi pedig generált állapot, amelyek mindegyike valamilyen szabály alkalmazásával jön létre az előző sorban lévő

állapotra.

Nézzünk egy példát: tegyük fel, hogy a [STR] rendszerben az $\square\square$ állapot levezetését szeretnénk. Íme egy levezetés, amely elvégzi a feladatot:

1. $\blacksquare\blacksquare$ kezdeti állapot
2. $\blacksquare\blacksquare\blacksquare$ [R2] "f" = $\blacksquare\blacksquare\blacksquare$ $\langle I = \emptyset$
3. $\cdot\square\square$ [R1] 'f' = $\blacksquare\blacksquare\blacksquare$
4. $\cdot\blacksquare\blacksquare\blacksquare\blacksquare\blacksquare$ [R2] "f" = $\square\square\langle I = \emptyset$
5. $\cdot\square\square\square\square$ [R2] 'f' = $\blacksquare\blacksquare\langle I = \blacksquare\blacksquare\blacksquare$
6. $\cdot\square\square\square$ [R1] 'f' = $\blacksquare\blacksquare\blacksquare$

A jobb oldali megjegyzések elárulják, hogy az állapotot hogyan kaptuk meg - melyik szabályt használtuk, és hogy a szabályban szereplő karakterláncváltozókat hogyan *instanciáltuk*. Így például a 2. állapotot az 1. állapotból (a kezdeti állapotból) úgy vezettük le, hogy az [R2]-t alkalmaztuk rá, és az 'f'-et $\blacksquare\blacksquare$ -nek, az $\langle I$ -t pedig \emptyset -nak (üres) tekintettük, ami egy $\blacksquare\blacksquare$ 'f' formájú instanciát eredményezett (ahol az 'f' $\blacksquare\blacksquare\blacksquare$), nevezetesen a $\blacksquare\blacksquare\blacksquare$.

Megjegyzendő azonban, hogy az [R2]-t a kezdeti állapotra másképp is alkalmazhattuk volna. Vehettük volna "f"-nek $\blacksquare\blacksquare$ -et és $\langle I$ -nek , ami a

$\square\square\square\square$ - τ eredményezte volna. Ez azt jelenti, hogy az [STR] egy *nemdeterminisztikus* formális rendszer. Egy rendszer akkor

determinisztikus, ha bármely adott állapotra legfeljebb egy szabály vonatkozik, és csak egyféleképpen. Ha egynél több szabály vonatkozik a rendszer bármelyik adott állapotára, vagy ha egy szabály egynél több módon vonatkozik a rendszer egy adott állapotára, akkor a rendszer

nem

determinisztikus.

7.2. gyakorlat

Hányféleképpen lehet [R2]-t alkalmazni az államra?

$\square\square\square\square$ és mi lenne a kimenet minden esetben? Mi a helyzet a $\blacksquare\blacksquare$ és $\square\square\square$ állapotokkal? Mi a minta?

Talán észrevetted, hogy az [R2] kezdeti állapotra való alkalmazásának eredménye az általunk tárgyalt alternatív módon (az "f"-et $\blacksquare\blacksquare$ helyett $\blacksquare\blacksquare\blacksquare$ -nek véve) megegyezik a levezetés 5. állapotával. Ez azt jelenti, hogy a példában a kezdeti állapottól valójában csak egy lépésben (négy helyett) juthatunk el az 5. állapotig.

Gyakran előfordul, hogy egy adott állapotot többféleképpen is le lehet vezetni. Gyakran a legegyszerűbb (azaz legrövidebb) levezetés megtalálása érdekel bennünket.

7.3. gyakorlat

Adja meg az $\square\square$ állapot legrövidebb levezetését [STR].

Elgondolkodhatunk azon - tekintettel az [STR] határozatlanságára és a levezetések pluralitására -, hogy van-e általában véve egy effektív eljárás a levezetések megtalálására egy formális rendszerben. Ez a klasszikus Mesterséges Intelligencia kutatási hagyomány számára nagyon fontos kérdésnek bizonyul, és a 11. fejezet elején erre fogunk összpontosítani. Egyelőre néhány informális heurisztika szolgál majd útmutatásul a következő feladatok során. Először is, dolgozzunk visszafelé a megoldásból. Vizsgáljuk meg a levezetni kívánt állapotot, és állapítsuk meg, hogy az lehetett-e bármelyik szabály kimenete. Ha nem, akkor nem kell folytatni; ha igen, akkor lesz egy útmutatója, hogy melyik bemenet(ek) adhatta(k) ezt a kimenetet. Próbáljon meg visszamenni a kiindulási állapotba, visszamunkázva-szabályokon keresztül.

Másodszor, törekedjen a céljára. Ha választhat a szabályalkalmazások közül egy olyan állapothoz, amely hosszabb, mint a célállapot, és az egyik választás rövidebb kimenetet eredményez, akkor valószínűleg (de nem garantáltan) ez a jó út.

7.4. gyakorlat

(a) A [STR] rendszert egészítsük ki a következő

szabállyal: [R3] ■ 'f₃ ■ 'f'f'f₃

és adjon megfelelően kommentált levezetéseket az állapotokhoz:

1	.□□ 4	.□
2	. 5	.□□□□□
3	.□ 6	.□

(b) Tudsz generálni egy állapotot anélkül, hogy ?
Magyarázza meg az érvelését.

7.5 GENERÁCIÓS FÁK

A 7.4(b) feladatra adott válasz rávilágít a rendszer [STR] egyik fontos tulajdonságára, nevezetesen, hogy nincsenek *végállapotai*.

Egy állapotot *terminálisnak* nevezünk, ha a rendszer olyan *generált* állapota, amelyre nem vonatkoznak szabályok. Tehát, bár ■■ az [STR] olyan állapota, amelyre nem vonatkoznak szabályok, mégsem *terminális* állapot, mivel nem generált állapot.

Az [STR] minden generált állapota *ipso facto* egy szabály kimenete, és ezért tartalmaznia kell egy . Mivel az [R2] minden olyan állapotra vonatkozik, amely bárhol tartalmaz a-t, nincsenek olyan generált állapotok, amelyekre nem vonatkoznak szabályok. Ezért az [STR]-nek nincsenek végállapotai.

Tekintsük a [BIN] formális rendszer alábbi specifikációját.

Vannak-e ennek a rendszernek végállapotai?

[BIN]

[S1]øegy állapot

[S2]Ha 'f' egy állapot, akkor 'f'1, 'f'0 és 'f'ro is.

[S3]kezdetiállapot: ro

[R1] 'f'ro <I əf' 1 <I / 'f' 0 <I / 'f'ro ro ro <I

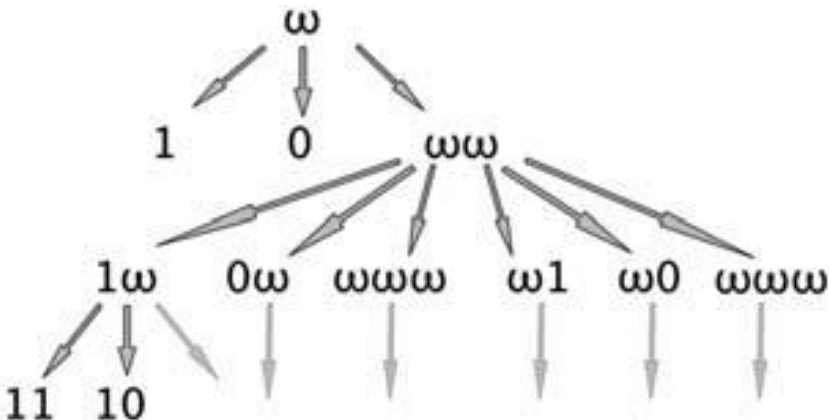
A rendszer egyetlen szabálya három lehetséges kimenet közül választhat - a '/' a 'vagy'-ot jelöli. Informálisan olvasva azt mondja, hogy bármelyik állapot ro átírható 1-nek, 0-nak vagy két ro-nak.

A kezdeti állapotot tekintve a szabály lehetővé teszi az 1, 0 és ro szimbólumok összes és csakis véges karakterláncának levezetését. Más szóval, a rendszernek nincs olyan állapota, amely ne lenne generálva. De vannak-e végállapotok?

Tekintsük meg, hogy a szabályt háromféleképpen alkalmazhatjuk a kezdeti állapotra. Az így kapott kimeneti állapotok 1, 0 és roro lesznek. Ezek közül az első kettő nem tartalmazza a ro szimbólumot. Mivel a rendszer egyetlen szabálya csak azokra az állapotokra vonatkozik, amelyek tartalmazzák a ro szimbólum legalább egy előfordulását, az 1 és 0 állapotok *végállapotok*.

Valójában egy kis gondolkodással megmutathatjuk, hogy a [BIN] végállapotai az 1 és/vagy 0 szimbólumok véges karakterláncai lesznek. Más szóval, a [BIN] végállapotai a bináris szimbólumok véges karakterláncai (pl. 1001011). Tehát bármely véges bináris karakterláncnak van egy levezetése a [BIN]-ben.

Ez világosabbá válik, ha a 7.3. ábrán látható módon felrajzoljuk a *generációs fát*. Fordítsuk meg az ábrát fejjel lefelé, és egy kicsit világosabbá válik, hogy miért hívják ezeket *fa* struktúráknak. A tetején lévő állapot a *gyökércsomópont*.



7.3 ábra Generációs fa.

Egy *csomópontból* induló nyíl egy szabály lehetséges alkalmazását jelenti az adott csomópont állapotára - a szabály alkalmazásából eredő kimeneti állapot a csomópontban van, ahová a nyíl vezet. A gyökércsomópontból tehát három nyíl vezet ki, mivel a szabály háromféleképpen alkalmazható rá. Ezeket a nyilakat a fa metaforának megfelelően *ágaknak* fogjuk nevezni.

A 7.3. ábra néhány elágazása *terminális csomópontokhoz* vezet. Egy csomópont akkor számít *terminálisnak*, ha nem vezetnek ki belőle nyilak (azaz ha a csomópontban terminális állapot van). A terminális csomópontok a fánk levelei vagy csúcsai.

A rendszer első *iterációját* a generációs fa első szintje képviseli a gyökércsomópont után. A 7.3. ábra az első két *iteráció* teljes generációs fáját és a harmadik *iteráció* részleges fáját mutatja.

Az első iteráció három csomópontot tartalmaz, amelyek közül kettő terminális. Összesen az első iteráció két ro-t tartalmaz - a kettő az egyetlen nem terminális csomópontnál. Tudjuk, hogy minden egyes ro-ra háromféleképpen lehet alkalmazni a szabályt, így az első iterációból hat ág vezet. Más szóval, a második iterációnál hat csomópont lesz (ahogy a példában látható).

7.5. gyakorlat

- A 7.3. ábrán látható fa harmadik iterációja nem teljes. A fenti bekezdésben leírtak alapján határozzuk meg, hogy hány csomópontnak kell lennie a harmadik iterációnál.
- Hány ilyen csomópont lesz terminál?
- Keress egy nagy papírlapot vagy egy táblát, és rajzold fel a [BIN] teljes generációs fáját a negyedik iterációig.

Egy generációs fát tekintve, a levezetések egyenesen kiolvashatók belőle, ha egyszerűen követünk egy ágot a minket érdeklő csomóponttól vissza a gyökércsomópontig. Itt van például a [BIN] "11" állapotának levezetése a 7.3. ábrán látható fa legbaloldali ágáról leolvasva.

1.ro	kezdeti	
	állapot	
2.roro	'f' = \emptyset	<I = \emptyset
3.1ro	'f' = \emptyset	<I = ro
4.11	'f' = 1	<I = \emptyset

Van tehát egy általános eljárás a levezetések megtalálására egy formális rendszerben: egyszerűen töltsük ki a rendszer teljes generációs fáját, keressük meg azt az állapotot, amelyhez levezetésre van szükségünk, és olvassuk ki a levezetést a fából.

Sajnos, ahogyan azt a 7.5(c) feladatnál láthattad, a generációs fák nagyon gyorsan nagyon bonyolulttá válhatnak. Sok rendszer, köztük a [BIN] is, *exponenciális robbanástól* szenved - az ágak exponenciálisan szaporodnak, ahogy iterálunk.

A legrosszabb az, hogy - amint arra már talán rájöttél - egyes ágak örökké tartanak - a végtelenségig tartanak. Ez azt jelenti, hogy a generációs fa felállításának eljárása nem *effektív olyan* rendszerek esetében, amelyek fája végtelen ágakkal rendelkezik. Ennek következményeit a 11. fejezetben a keresési eljárások tárgyalásakor fogjuk tovább vizsgálni.

7.6. gyakorlat

Adja meg a [BIN] levezetéseit az állapotokra:

- (a) 1001
- (b) 0100101
- (c) 000111

7.6 FORMALITÁS ÉS IZOMORFIZMUS

Van még egy utolsó megjegyzés a formális rendszerekkel kapcsolatban, mielőtt továbblépnénk, és valami érdekesebbet kezdenénk velük. Fontos megérteni, hogy a formális rendszerek egyetlen fontos vagy releváns tulajdonsága a *formális* tulajdonságok - a *forma* tulajdonságai.

Egy formális rendszer működése szempontjából soha nem fontos, hogy a rendszer fizikailag hogyan valósul meg. Vegyük újra a sakkot. A bábuk lehetnek fából faragott vagy kőből faragott figurák, lehetnek papíron vagy elektronikus képernyőn megjelenő szimbólumok, lehetnek érmék vagy más, rögtönzött szolgálatba állított zsetonok, de akár emberek is lehetnek egy megfelelően nagy táblán.

Az állapotok megkülönböztetése és a szabályok alkalmazása szempontjából csak a rendszer effektíven megkülönböztethető *formákba* rendeződései relevánsak.

Másképpen fogalmazva, egy formális rendszer műveletei teljesen függetlenek attól a közegtől (vagy szubsztráttól), amelyben *instanciálódnak*.

Ennek emlékeztetnie kell Önt a funkionalista elméletek által állított szubsztrátfüggetlenségre. Egy funkionalista számára az egyetlen lényeges dolog, amit a mentális állapotokról tudni kell, azok a *funkciók*. Hasonlóképpen, amikor formális rendszereket vizsgálunk, az egyetlen releváns dolog, amit tudnunk kell, a *formák*.

Egy formális rendszer műveletei teljesen függetlenek a rendszer bármely értelmezésétől. Bár a formális rendszerek elvileg értelmezhetőek (például egy sor instanciált formális rendszert értelmezhetek sakkjátszma-ként), nem kell semmilyen értelmező munkát végezni ahhoz, hogy szabályokat alkalmazzak az állapotokra - csupán algoritmikus eljárásokat kell követni.

Tehát, ahogyan az már valószínűleg nyilvánvaló, ha egy [A] rendszert vizsgálok, amely rendelkezik valamilyen [B] rendszer összes és csakis ugyanazon formális tulajdonságával, akkor *éppen a* [B] rendszert vizsgálom. Ha két rendszer *formailag ekvivalens*, akkor *ugyanannak a rendszernek* a példányai. Akár bábukkal, szimbólumokkal, érmékkel vagy emberekkel sakkozom, sakkozom.

Ha két rendszer formailag ekvivalens - ha minden és csakis ugyanazokkal a formális tulajdonságokkal rendelkezik -, akkor azt mondjuk, hogy *izomorfikusak* egymáshoz, vagy ugyanannak a formális rendszernek *az izomorfizmusai*.

Egy [A] formális rendszer *izomorfikus* egy [B] formális rendszerrel, iff a szimbólumok egységes helyettesítésével [B]-t levezethetjük [A]-ból. Vegyük például az alábbiakban megadott rendszert:

[S1] Ø egy állapot
 [S2] Ha X egy állapot, akkor Xa és
 Xb is [S3] Kezdeti állapot: *aabb*
 [R1] *aaXb bX*
 [R2] *XbYaaX*
b

ahol X és Y karakterláncváltozók

Elég világosnak kell lennie, hogy a fenti példa izomorfikus az [STR] eredeti bemutatásával. Valójában vegyünk bármilyen tetszőleges szimbólumot, és helyettesítsük egységesen *a-val*, és hasonlóképpen *b-vel*, és az eredmény egy másik [STR] izomorfizmus lesz. A "szimbólum" kifejezés itt elég tágan értelmezhető, és magában foglalja a fizikai jeleket, például érméket vagy embereket - használhatnánk például férfiak és nők rendezett sorát az [STR] vizsgálatához (feltéve, hogy effektíven meg tudjuk őket különböztetni).

A lényeg itt az, hogy bármilyen formális rendszerre, amelyet vizsgálni szeretnénk, létezik egy izomorf szimbólumrendszer. Ez jó hír, ha automatizált módszereket szeretnénk alkalmazni a formális rendszerek vizsgálatára.

80

Most, hogy a formális rendszereket, azok jellemzőit és működését jól ismerjük, itt az ideje, hogy a formális rendszereket a gyakorlatban is alkalmazzuk.

felhasználás, amelyre bevezettük őket. A következő fejezetben megnézzük, hogyan használhatunk egy bizonyos fajta formális rendszert a *kom-putációhoz*.

Hadd mondjam el még egyszer, hogy mielőtt folytatnátok, fontos, hogy elsajátítsátok az anyagot eddig a pontig. A következő két fejezetben erre az alapra fogunk tovább építeni, így ha van valami, amiben bizonytalanok vagytok, most van itt az ideje, hogy átdolgozzátok. Ha viszont készen állsz egy nagyobb kihívást jelentő anyagra, akkor olvass tovább.

8. FEJEZET

COMPUTABILITY

Az eddig vizsgált formális rendszerek nagyon kezdetleges húrrendszerek voltak. Következésképpen hasznos alkalmazásuk meglehetősen korlátozott. A formális rendszereket azonban sokkal érdekesebb és hasznosabb célokra is alkalmazhatjuk. Különösen a formális rendszereket használhatjuk *számítások* elvégzésére. Ebben a fejezetben a determinisztikus formális rendszerek egy bizonyos fajtáját - a *regisztergépet* - fogjuk használni a *kiszámíthatóság* szigorú definiálására.

8.1 PÉNZTÁRGÉPEK

A regisztergépek elméleti entitások. Fizikailag azonban megvalósíthatók (mint bármely formális rendszer). Az általunk ismert modern digitális számítógépek a regisztergépek egy speciális fajtájának megvalósításai, amint azt a következő fejezetben látni fogjuk.

Az egyszerű regisztrológépek, mint amilyeneket ebben a fejezetben vizsgálunk, egyszerűen megvalósíthatók kőhalmokkal, érmékkel vagy más fizikai zsetonokkal. Ha van kéznél egy érme-vagy más zsetongyűjtemény, akkor nagy valószínűséggel hasznos lesz, ha ezeket magadnál tartod, amikor a fejezetben szereplő példákön és gyakorlatokon dolgozol. A galambdúc analógia nagyon találó, amikor először kezdünk el gondolkodni a pénztárgépekről. Emlékezzünk vissza a 7.2. szakaszból, hogy az általunk választott entitások bármilyen gyűjteménye felett definiálhatunk állapotokat, feltéve, hogy két adott állapot effektíven megkülönböztethető. Az állapotokat különösen definiálhatjuk úgy, mint a betűk eloszlását a galambdúcbokban. Ez közvetlenül analóg azzal, ahogyan a regisztergépek állapotait szeretnénk definiálni.

A regisztergépek állapotai a *regiszterek* véges sorozatának tartalma. A regiszterek diszkrét tárolóként értelmezendők, ezért a galambdúc analógia. Ezeket a regisztereket R_0, R_1, R_2, \dots , stb. A regiszterek száma végtelen lehet; mi azonban azt az egyszerűsítő

feltételezést tesszük, hogy csak véges számú regiszter tartalma van egy adott időpontban.

A regiszter *tartalma* természetes számként ábrázolható - a regiszterben található elemek száma. Ha tehát van egy galambdobozunk három betűvel, vagy egy diszkrét halom három érmével, akkor azt mondhatjuk, hogy van egy regiszterünk, amely *három* (elemet) tartalmaz. Az, hogy pontosan mik a tételek (betűk, érmék, válogatott tárgyak stb.), egyáltalán nem számít. Bármely olyan entitás, amely effektíven megkülönböztethető állapotokat képes alkotni, szolgálhat egy formális rendszerben manipulált szimbólumként.

Három galambdúc sorozat, amely egy, három, illetve két betűt tartalmaz, izomorf lesz az egy, három, illetve két érmét tartalmazó kupacok sorozatával. Mindkettő izomorf egy olyan regiszteres gépnek, amelynek tartalma *egy, három, kettő, az első három regiszterben van.*

A regisztergépek állapotai tehát természetes számok véges sorozataként ábrázolhatók. Az 1, 4, 16, 2, 27 sorozat egy olyan regisztergépet reprezentál, amelynek R0-ban 1, R1-ben 4, R2-ben 16, R3-ban 2 és 27 van.

R₄ . Más szóval, a számsorozat egy rendezett

öt, 1, 4, 16, 16, 2 és 27 tárgyat tartalmazó halom sorozata.

8.2 PROGRAMOK

A nyilvántartó gépek formális rendszerek. Most már tudjuk, hogy mik a regisztergépek *állapotai* - minden értelemben és célra egyszerűen természetes számok véges sorozatairól van szó. Ezután tudnunk kell, hogy mik a *szabályok*.

A regisztergépeknek csak egy szabálya van. Ez a szabály egy speciális formát ölt, amelyet *programnak* nevezünk.

A regiszteres gépi *program* véges számú sorból áll, amelyek mindegyike két összetevőből áll: egy *sorszám*ból és egy *utasítás*ból. A sorszámok egyszerűen természetes számok, amelyeket a program soraira való hivatkozás megkönnyítésére rendelnek hozzá. Minden sorhoz egyedi sorszámot kell rendelni. Ezek a sorszámok az olvashatóság érdekében szokásosan egymást követő sorszámok, de nem feltétlenül azok. Az *utasítások* kétféle formát ölthetnek: ezek vagy *növelő utasítások*, vagy *csökkentő utasítások*.

Az *inkrementálási utasítás* az $I a b$ formájú, ahol az I a "inkrementálást" jelenti, az a és b pedig numerikus változók - természetes számok. A *dekrement utasítás* a $D a b c$ alakú, ahol a D a "decrement"-et jelenti, és a kisbetűs latin betűk a növekmény utasításhoz hasonlóan numerikus változókat jelentenek.

Emlékezzünk vissza, hogy effektív eljárásoknak kell létezniük mind annak meghatározására, hogy egy szabály vonatkozik-e egy formális rendszer állapotára, mind pedig a szabályok alkalmazására

a formális rendszerek állapotaira.

Az effektív eljárás annak meghatározására, hogy egy regisztergép-program egy adott regisztergép-állapotról vonatkozik-e, a következő:

1. Vizsgáljuk meg az R_0 tartalmát (ez az első regiszter mindig erre a célra van félretéve, és *programszámlálónak* vagy *pc-nek* nevezzük).
2. Ha a programnak nincs olyan sora, amely a pc-ben lévő számmal kezdődik, akkor a program nem vonatkozik az államra.
3. Ha a programban van olyan sor, amely a pc-ben lévő számmal kezdődik, akkor a program az adott államra vonatkozik.

A regiszteres gépi program alkalmazásának effektív eljárása a következő:

1. Ha a programnak az a sora, amely a pc-ben lévő számmal kezdődik, tartalmaz egy *inkrementáló utasítást* ($I\ a\ b$), akkor *növeljük* (adjunk hozzá egyet) R_{a-t} és tegyük $b-t$ a pc-be.
2. Ha a programnak az a sora, amely a pc-ben szereplő számmal kezdődik tartalmaz egy *dekrementáló utasítást* ($D\ a\ b\ c$), majd *dekrementál* (kivesz egyet) $R_{a-ból}$ és $b-t$ tesz a pc-be. Ha R_a már üres (nullát tartalmaz), akkor ne tegyünk semmit, csak tegyük $c-t$ a pc-be.

Az $I\ a\ b$ és a $D\ a\ b\ c$ utasításformák pontosan ezek: *formák*. A regisztergépes utasítások ezeknek a formáknak a *példányai*. Az instanciált formák értékeket rendelnek a változókhoz - ebben az esetben természetes számokat a numerikus változókhoz. Így például egy regisztergépes program tényleges sorai így néznek ki:

```
1 I 1 2
2 D 3 4 3
```

Az 1. sor utasít bennünket, hogy növeljük az R_1 értéket, majd tegyünk 2-t a pc-be (R_0). A 2. sor arra utasít minket, hogy csökkentsük az R_3 értéket, *ha tudjuk*, akkor tegyünk 4-et a pc-be, egyébként (ha az R_3 üres) csak 3-at tegyünk a pc-be.

Ha tehát végignézzük egy regiszteres gépi program sorait, az utasítás betűi utáni első oszlopban lévő számok ~~regiszter~~ mutatnak - ezek mondják meg, hogy melyik regisztert kell növelni vagy csökkenteni. A következő oszlopban lévő számok megmondják, hogy az utasítás sikeres végrehajtása után melyik számot kell a pc-be helyezni. A harmadik oszlopban lévő számok (amelyek csak a dekrementáló utasításokban jelennek meg) megmondják, hogy melyik számot kell a pc-be helyezni, ha nem tudjuk végrehajtani az utasítást.

A regiszterek mindig növelhetők, mivel nincs legnagyobb természetes szám. Nem lehet azonban mindig dekrementálni őket. A regiszterek tartalma *természetes számok*, nem *egész számok*. A

74

természetes számok közé tartozik a nulla és a *pozitív* egész számok.
Ha egy regiszter 1-t tartalmaz, akkor

még egyszer lecsökkentjük, akkor azt mondjuk, hogy a regiszter üres. Egy üres regisztert nem lehet dekrementálni. Egy halom semmiből nem lehet elvenni valamit.

A program alkalmazhatóságának meghatározására és a program alkalmazására szolgáló effektív eljárásokat egyetlen algoritmussá egyesíthetjük a program *futtatására*, a következőképpen:

1. Nézd meg a számítógépet.
2. Ha nincs olyan programsor, amely ezzel a számmal kezdődik, akkor álljon meg.
3. Ha van egy programsor, amely ezzel a számmal kezdődik, akkor hajtsa végre az abban a sorban lévő utasítást.
4. Ismételd meg.

Egy program *futtatása* tehát egyszerűen abból áll, hogy először meghatározzuk, hogy a program vonatkozik-e az aktuális állapotra, majd végrehajtjuk a megfelelő utasítást, amíg a program már nem alkalmazható (amíg nem lesz olyan állapot, amelynek a pc-ben olyan szám van, amelynek nincs megfelelő programsor).

8.3 EGY PROGRAM FUTTATÁSA

Most, hogy már tudjuk, hogy mik a regisztergép állapotai, és hogy mit jelent egy regisztergépes program *futtatása*, vizsgáljuk meg egy egyszerű regisztergép műveleteit. Tekintsük a következő programot:

[ADD]

1 D 2 2 3
2 I 1 1

Példázzuk ennek az egyszerű programnak a működését, és próbáljuk meg meghatározni, hogy mit csinál. Az utasítás betűi utáni első oszlopból tudjuk, hogy ebben a programban csak két regiszterre hivatkoznak - R_1 és R_2 . Alkalmazzuk tehát ezt a programot egy olyan kezdeti állapotra, amelynek tartalma ebben a két regiszterben van, mondjuk 3 az R_1 -ben és 2 az R_2 -ban. Tegyük fel azt is, hogy a kezdeti állapotunk az R_0 -ben (a pc) 1-et tartalmaz - ez hagyományos, mivel ez biztosítja, hogy először a programunk 1. sorának végrehajtása fog történni.

Ezt a kezdeti állapotot a 8.1. ábra első sora mutatja. A kezdeti állapot alatti minden egyes sor a program egy-egy alkalmazásának eredő állapotát ábrázolja.

Ha az [ADD] programot a leírt kezdeti állapottal *futtatjuk*, a műveletek sorrendje a 8.1. ábrán látható lesz.

R_0	R_1	R_2
1	3	2
2	3	1
1	4	1
2	4	0
1	5	0
3	5	0

8.1. ábra A műveletek sorrendje.

Ha kéznél vannak a zsetonok, vedd elő őket, és rendezd el őket három kupacba, hogy a fenti példában a kezdeti állapotot ábrázolják. Most pedig dolgozzuk végig lépésről lépésre a példát. Egyszerűen a program futtatására szolgáló algoritmust fogjuk alkalmazni, amelyet a 8.2. szakasz végén ismertettem.

Először is megnézzük az R_0 (a pc) tartalmát a kezdeti állapotban. Ebben 1 van, és a programnak van egy sora, amely 1-gyel kezdődik, így az 1. sorban lévő utasítást hajtjuk végre. Az 1. sor egy dekrementáló utasítást tartalmaz - azt mondja, hogy dekrementáljuk az R_2 -t (ami 1-et hagy az R_2 -ben), majd tegyünk 2-t a pc-be. Ennek az első alkalmazásnak az eredményül kapott állapotát a 8.1. ábrán a kezdeti állapot alatti sor ábrázolja. Hogy egyértelmű legyen, most 2 van a pc-ben, 3 az R_1 -ban és 1 az R_2 -ben.

Most megismételjük a folyamatot. A pc tartalmaz 2-t, és van egy 2-es sorunk, így végrehajtjuk ezt az utasítást. A 2. sorban lévő utasítás egy növelő utasítás, amely azt mondja, hogy növeljük R_1 (ami 4-et tesz R_1), majd tegyünk 1-et a pc-be. Az eredményül kapott állapotot a 8.1. ábra harmadik sora mutatja.

A pc most 1-et tartalmaz, így újra végrehajtjuk az 1. sorban lévő utasítást, amely dekrementálja az R_2 értéket (így az R_2 értékben 0 marad), és 2-t tesz a pc-be. Ezután ismét végrehajtjuk a 2. sorban lévő utasítást (mivel most már 2 van a pc-ben), inkrementáljuk az R_1 értéket (így az R_1 érték 5 lesz), és 1 kerül a pc-be.

Most már 1 van a pc-ben, 5 az R -ben₁ és 0 az R -ben₂. Amikor megpróbáljuk újra végrehajtani az 1. sorban lévő utasítást, azt látjuk, hogy nem tudjuk. Az R_2 üres, tehát nem lehet dekrementálni. Ehelyett csak 3-at teszünk a pc-be.

Amikor aztán újra alkalmazni kezdjük a programot, azt látjuk, hogy az már nem alkalmazható. Van egy szám a pc-ben, amelynek nincs megfelelő programsor, így a program megáll. A végállapot, ahová eljutunk, olyan, amelyben az R_1 -ben 5 van, és amelyben az R_2 üres.

8.1. gyakorlat

- (a) Futtassa a [ADD] programot néhányszor differenciális kiindulási állapotokkal. Minden kezdeti állapotnak 1-et kell tartalmaznia a pc-ben, de az R_1 és az R_2 értékeket tetszés szerint választhatjuk meg. Használjunk ceruzát és papírt, ha csak ez áll rendelkezésünkre, de úgy találhatjuk, hogy nagy segítséget jelent, ha egy halom tokent használunk.
- (b) A program mindig úgy fejeződik be, hogy az R_1 értéket kap. Hasonlítsuk össze ezt minden esetben azokkal az értékekkel, amelyeket R_1 és R_2 a kezdeti állapotban vett fel. Meg tudod állapítani, hogy mit csinál a program?

8.4 SZÁMÍTÁS

A 8.1. feladat elvégzése után észrevehettük, hogy az [ADD] program mindig egy olyan számmal végződik az R_1 számmal, amely az R_1 és az R_2 számok összege a kezdeti állapotban. Más szóval, az [ADD] program összeadja az R_1 és az R_2 számokat, és azok összegével végződik az R_1 számban.

A program ezt úgy éri el, hogy egy nagyon egyszerű összeadási algoritmust valósít meg. Ha van két halom összeadandó dolog, akkor függetlenül attól, hogy hány darab van a két halomban, vagy

melyik a nagyobb, a következő effektív eljárás végrehajtásával meg tudjuk találni az összegüket. Először próbáljunk meg elvenni egyet az egyik halomból. Ha ez sikerül, adjuk hozzá

az egyiket a másik kupacba. Ezt addig ismételjük, amíg az első kupac ki nem ürül, ekkor a két kupac összege a második kupacban lesz. Az [ADD] program pontosan ezt teszi (ahol az "első halom" az R_2 , a "második halom" pedig az R_1).

Adott két tetszőleges x és y szám, az [ADD] program létrehozza az összegüket. Másképpen úgy is mondhatjuk, hogy az [ADD] program *kiszámítja* az összeadást.

Az előző bekezdésben jelent meg először az a fogalom, amelynek kifejtése e fejezet célja, nevezetesen a *számítás*. Úton vagyunk annak pontos formális leírása felé, hogy mit jelent *számolni*. Először azonban van néhány dolog, amit *informálisan* mondhatunk a számításról.

Egy regisztergép program és egy kezdeti állapot bármely adott kombinációjára a program alkalmazásainak sorozata egy *komputáció*.

Ezt teszik a nyilvántartó gépek - *számolnak*. Hogy pontosan mit is számolnak, arra mindjárt rátérünk.

Mielőtt továbblépnénk, megjegyezhetjük, hogy vannak olyan regiszteres gépi programok, amelyek soha nem fejeződnek be. Ennélfogva vannak olyan *számítások*, amelyek soha nem érnek véget. Tekintsük az egysoros regisztergép programot, amely 1 1 1 1 1 olvas. Ez a program soha nem fog ~~lefejező~~ *lefejező*, egyszerűen a végtelenségig folytatja az R_1 inkrementálását.

Sajnos nincs hatékony eljárás annak megállapítására, hogy egy adott program leáll-e vagy sem. Ez a számítógépes programozók által jól ismert *megállási probléma*. Vannak bizonyos buktatók, amiktől óvakodni kell a programozásban, különösen az egymásba ágyazott ciklusok esetében, amelyek ahhoz vezetnek, hogy a program soha nem ér véget. A programozókat arra tanítják, hogy erre figyeljenek, és vannak módszerek és konvenciók a nem véges ciklusok elkerülésére. Nincs azonban olyan *algoritmus*, amellyel ellenőrizhetnénk, hogy egy adott program véget ér-e. Természetesen léteznek olyan szoftverellenőrző eszközök, amelyek többek között ismétlődő mintákat keresnek a *valószínű* nem-megszűnés bizonyítékaként. Vannak azonban olyan módszerek a végtelenítésre, amelyek nem ismétlődéssel járnak - gondoljunk csak a természetes számok sorozatára vagy az r számjegyeire.

8.5 KISZÁMÍTHATÓ FÜGGVÉNYEK

A *számítást* eddig informálisan egy regiszteres gépi program műveleteiként mutattuk be. Láttunk egy példát arra is, ami nem *számítható*. Nincs olyan effektív eljárás (és még inkább nincs olyan regiszteres gépi program), amely meghatározná

hogy egy adott program leáll-e. Következésképpen valami, amit nem lehet kiszámítani, az a kérdés, hogy egy adott számítás meg fog-e állni.

Pontosan milyen *dolgok* számíthatók? E kérdés megválaszolásához a *kiszámíthatóság* fogalmának formális jellemzésére van szükség, ami viszont egy kis matematikai terminológiát igényel.

Szükségünk van a *függvény* formális fogalmára. Ezt a fogalmat mindenki ismeri az általános iskolai számtanból.

Az $f(x_1, x_2, \dots, x_n) = m$ *függvény* egy matematikai összefüggés egy meghatározott számú n bemenet és egy egyedi kimenet m .

Az összeadás egy *függvény*. Fix számú bemenetet (kettőt) fogad el, és minden adott bemeneti párhoz egyedi kimenet tartozik. A szorzás, a szubtrakció, az osztás, a hatványozás és az összes többi alapvető számtani művelet szintén *függvény*.

Egyes függvények bizonyos bemenetek esetén nem definiáltak. Ezeket a függvényeket *részleges függvényeknek* nevezzük, ellentétben a *teljes függvényekkel*, amelyek kimenete minden lehetséges (megfelelő típusú) bemenetre definiált.

Az összeadás és a szorzás *totális függvények*. A bemeneti típusuk valós számok, a kimenetük pedig valós számok bármely lehetséges párjára definiált.

Az osztás viszont egy *részleges függvény*. Ez is valós számokat fogad el bemenetként, de a kimenete bizonyos bemeneti párok esetében nem definiált (a 0-val való osztás nem definiált).

A függvény adicitásának nevezzük azt, hogy egy függvény hány n bemenetet vesz fel. Az összeadás és a szorzás adicitása 2. A négyzetképző függvény adicitása 1.

Az *adicity* főnévnek is vannak melléknévi rokonai megfelelő numerikus előtaggal. Azt mondanánk például, hogy a négyzetelés *monádikus* függvény, az összeadás és a szorzás pedig *dyádikus* függvény. Kevésbé technikailag ezeket *egyhelyes* függvényeknek, *kéthelyes* függvényeknek stb. is nevezhetjük.

Ismerjük az összes szükséges terminológiát ahhoz, hogy pontos formális definíciót adjunk a *számítás* fogalmának a következők szerint.

Ha f egy n -es adicitású függvény, akkor a P program *kiszámítja* az f ha:

amikor a P-t úgy futtatjuk, hogy a pc-ben $1, x_1, \dots, x_n$ az R_1 -ben, \dots, R_n és az összes többi regiszter üres, akkor:

- Ha $f(x_1, \dots, x_n)$ nem definiált, akkor a számítás soha nem ér véget.
- Ha $f(x_1, \dots, x_n) = m$, akkor a számítás azzal ér véget, hogy
 m
 az R_1 -ben.

Ez egy elég sűrű meghatározás, ezért bontsuk ki, és nézzünk meg néhány példát.

Az első dolog, amit meg kell jegyezni, hogy a számítás tárgyai a *funkciók*. A fenti definíció meghatározza azokat a feltételeket, amelyek mellett azt mondhatjuk, hogy egy adott regiszteres gépi program *kiszámít* egy függvényt.

A fogalom meghatározás a következőképpen hangzik. Először is van egy kezdeti feltétel, amely kimondja, hogy a programot úgy kell futtatni, hogy a pc -ben 1, a függvény n bemenete az első n regiszterben van, és az összes többi regiszter üres.

Így például annak meghatározásához, hogy az [ADD] program ~~the~~ a dyadikus összeadás $f(x_1, x_2) = x_1 + x_2$ függvényét, először is teljesítjük ezt a kezdeti feltételt az összes regiszter törlésével és a pc -be 1, az R_1 -be x_1 , az R_2 -be x_2 kerül. Ezután futtatjuk a programot.

Mivel az összeadás egy totális függvény, tudjuk, hogy a kimenete a következő lesz

meghatározott. Vagyis lesz egy m , amely a függvény komputálásának eredménye. Ahhoz tehát, hogy az [ADD] programról azt mondhassuk, hogy kiszámítja az összeadási függvényt, a fenti definíciónk szerint az R_1 -ben $x_1 + x_2$ -vel kell végződnie.

Az [ADD] program valójában mindig az $x_1 + x_2$ értékkel végződik.
 R_1 , ha a meghatározott inisialis feltételek teljesülnek.

Következésképpen most azt mondhatjuk (a formális feltételek bizonyított teljesülése alapján), hogy az [ADD] program kiszámítja a kétértékű összeadási függvényt.

Most már rendelkezünk a számítás pontos formális leírásával, amelyet ki akartunk dolgozni. Ennek segítségével a következőképpen definiálhatjuk a *kiszámíthatóságot*.

Egy függvény *kiszámítható* iff létezik egy olyan regiszteres gépi program, amely kiszámítja.

Összefoglalva tehát, a regiszteres gépek műveletei számítások. A számítások tárgyai függvények - a regisztergépek függvényeket számolnak. Egy függvényt *kiszámítani* annyit jelent, mint a fentiekben meghatározott formai feltételeknek megfelelni. Azt mondani, hogy egy függvény *kiszámítható*, azt jelenti, hogy van legalább egy olyan regisztergép, amely kiszámítja azt.

8.6 ÉPÍTÉSI PROGRAMOK

A fejezet hátralévő részében olyan módszereket dolgozunk ki, amelyekkel algoritmusokat megvalósító regiszteres gépi programokat készíthetünk. Különböző függvények kiszámítására fogunk programokat írni. Minden esetben egy feladatot fogok kitűzni, majd egy lehetséges megoldást fogok kidolgozni. A feladatok megkísérlése a megoldás elolvasása előtt jelentősen segíti az anyag megértésének megszilárdítását.

8.2. gyakorlat

Írjunk egy olyan regisztergépes programot, amely az R_1 tartalmát R_2 -be és R_3 -ba is átmásolja, az R_1 -et üresen hagyva. Tegyük fel, hogy úgy kezdjük, hogy R_2 és R_3 egyaránt üres.

A regiszteres gépi programok írásának minden nehézsége az algoritmus meghatározásában rejlik. Ha már meghatároztuk azt a lépésenkénti folyamatot, amely garantáltan eredményt ad, akkor ennek lefordítása regisztergépes kódra meglehetősen egyszerű. Ezért hasznos a ~~l~~katokenek alkalmazása - ezek segítenek nekünk lépésről lépésre végigmenni a lehetséges algoritmikus eljárásokon.

A 8.2. feladat algoritmus a következő. Először próbáljunk meg elvenni egy követ az első kupacból. Ha sikerül, tegyünk egy követ a második kupacba és egy követ a harmadik kupacba. Amikor az első halom kiürül, a tartalma átmásolódik a második és a harmadik halomba is.

Ha ezt lefordítjuk regiszteres gépi kódra, akkor megkapjuk:

```
1D      124
        2I23
        3I31
```

8.3. gyakorlat

Írjon egy olyan regiszteres gépi programot, amely az R_1 tartalmát átmásolja az R_2 -be, de az R_1 tartalmát megőrzi. Tegyük fel, hogy úgy kezdjük, hogy R_2 és R_3 egyaránt üres.

Nagyon gyakran szeretnénk egy regiszter értékét egy másik regiszterbe másolni, de úgy, hogy az első regiszter értéke *ne* vesszen el. Mivel az érték másolásának egyetlen módja az, hogy az értéket tartalmazó regisztert dekrementáljuk, miközben a célregisztert inkrementáljuk, mindig elveszítjük az első regiszter tartalmát. Ugyanezt az eljárást használhatjuk az érték visszahelyezéséhez is

az első regiszterbe, de aztán elveszítjük a célregiszterből, és ott találjuk magunkat, ahonnan elindultunk.

Ezt a dilemmát úgy kerülhetjük meg, ha egy *munkaregisztert* használunk. Nagyon gyakran szükségünk lesz *munkaregiszterek* alkalmazására a programokban, hogy a számítás során értékeket tároljunk.

Ebben az esetben azt kell tennünk, hogy az R_1 tartalmát átmásoljuk R_{2-be} és $R3$ -ba, majd az $R3$ tartalmát visszamásoljuk az R_{1-be} . Az $R3$ a munkaregiszterünk - egyetlen funkciója, hogy a bejövő érték másolatát tárolja.

tially az R -ben₁, hogy vissza tudjuk tenni, miután befejeztük a másolást az R -be₂.

A 8.3. feladat megoldását egyszerűen kibővíthetjük a következőképpen:

	1	1	2
D		4	
	2I	2	3
	3I	3	1
	4	3	5
D		6	
	5I	1	4

A 8.2. és 8.3. feladat megoldásai szigorúan véve nem "másoló" függvények. Emlékezzünk vissza a függvény kiszámításának definíciójából, hogy a számítások mindig az R_1 -ban lévő eredménnyel végződnek. A 8.2. feladat megoldása tehát valójában azt a függvényt számítja ki, amely bármely bemenetet a 0 értékre - $f(x) = 0$ - és

a 8.3. feladat megoldása kiszámítja azt a függvényt, amely bármelyik bemenet önmagára - $f(x) = x$.

A mi érdeklődésünk azonban nem azokra a funkciókra irányul, amelyeket ezek a programok hanem inkább az általuk alkalmazott módszerekben. A veszteség nélküli másolás módszere gyakran szerepel majd a programjainkban. Az utolsó két gyakorlat lényege egyszerűen az volt, hogy egy hasznos kódrészletet fejlesszünk ki, amelyet alprogramként használhatunk további programokban.

A következő feladat megoldásához a veszteség nélküli másolás módszerét kell alkalmaznunk.

8.4. gyakorlat

Írj egy olyan regisztergépes programot, amely kiszámítja a szorzási függvényt: $f(x1, x2) = x1 \times x2$. Tegyük fel, hogy a pc-ben 1-gyel kezdünk, $x1$ az R_{1-ben} és $x2$ az R_{2-ben} van, de a többi regiszter tartalmára vonatkozóan nem teszünk

feltevéseket.

Mint mindig, most is az algoritmus meghatározása jelenti a nehézséget. Két szám x és y szorzása úgy értelmezhető, hogy x másolatot halmozunk fel az x

y . Tehát az algoritmus, amit meg akarunk valósítani, megpróbálja csökkenteni R_1

és ha sikeres, akkor az R_2 egy példányát hozzáadja egy munkaregiszterhez. Ha az R_1 üres, akkor a munkaregiszterben felhalmozott értéket (y x másolatát) az R_1 -be akarjuk áthelyezni, és befejezzük a műveletet.

Tudjuk, hogy szükségünk van legalább egy munkaregiszterre, amelyben felhalmozhatjuk az eredményt. Szükségünk lesz még egy másik munkaregiszterre is, hogy veszteség nélkül másolhassunk az R_2 -ből az ~~eredményt~~ felhalmozó regiszterbe. Vagyis minden egyes alkalommal, amikor sikeresen dekrementáljuk az R_1 -t, az R_2 -t egy megoldásakkumulátorba (mondjuk R_3), valamint egy másik munkaregiszterbe (mondjuk R_4) akarjuk másolni, majd az R_4 -t vissza akarjuk tenni az R_2 -be, mielőtt újra megpróbáljuk dekrementálni az R_1 -ot.

Azt mondták nekünk, hogy ne tegyünk semmilyen feltételezést a függvény bemeneteit tartalmazó R_1 és R_2 regisztereken kívül. Ez azt jelenti, hogy programunknak először is meg kell győződnie arról, hogy a munkaregiszterek üresek. Ez könnyen elvégezhető egy-egy programsorral regiszterenként.

Tehát az algoritmus, amelyet meg akarunk valósítani, a következőket fogja tenni. Először is gondoskodjunk arról, hogy a munkaregiszterek (R_3 és R_4) üresek legyenek. Kísérjük meg az R_1 csökkentését. Ha ez nem sikerül, akkor az R_3 -at (a megoldás akkumulátora) helyezük át az R_1 -be. és megáll. Ha sikeres, másoljuk R_2 -t R_3 -ba és R_4 -be, majd mozgassuk vissza R_4 -et.

az R_2 címre, mielőtt újra megpróbálná csökkenteni az R_1 értékét.

Ha ezt lefordítjuk regiszteres gépi kódra, a következőket kapjuk:

```

1D 3 1 2
2D 4 2 3
3D 1 4 9
4D 2 5 7
 5I 3 6
 6I 4 4
7D 4 8 3
 8I 2 7
9D 3 10 12
10 1 9

```

I

8.5. gyakorlat

Írjon egy olyan regiszteres gépi programot, amely kiszámítja a négyzetszámítási függvényt:

$f(x) = x^2$. Tegyük fel, hogy a pc-ben 1-gyel és az R-ben x -szel kezdünk₁, de nem teszünk feltevéseket más tartalmakra

86

vonatkozóan.
regiszterek.

A négyzetelés egyszerűen a szorzás egy speciális esete - az az eset, amikor mindkét szorzandó azonos. Tehát x^2 egyszerűen x másolata x -nek.

A négyzetes függvény kiszámítására szolgáló regiszteres gépi program tervezésének egyik módja az lenne, ha a 8.4. feladat megoldásához használt szorzási algoritmuson kisebb változtatásokat végeznénk a következők szerint. Az R3 és R4 törlése mellett az R_{2-t} is töröljük. Ezután, mielőtt megkísérelnénk az R_1 dekrementálását, bemásoljuk az R_1 -et az R_2 -be, az R_3 -at használva munkaregiszterként a veszteség nélküli másoláshoz. Ezután követjük az R_1 és R_2 szorzásának algoritmusát.

Van azonban egy másik algoritmus is, amelyet megvalósíthatunk. Gondoljunk a következő regiszteres gépi programra:

```

1D 2 1 2
2D 3 2 3
3D 4 3 4
4I 2 5
5D 1 6 13
6D 2 7 9
7I 3 8
8I 4 6
9I 4 10
10 4 11
I
11 4 12 5
D
12 2 11
I
13 3 14 15
D
14 1 13
I

```

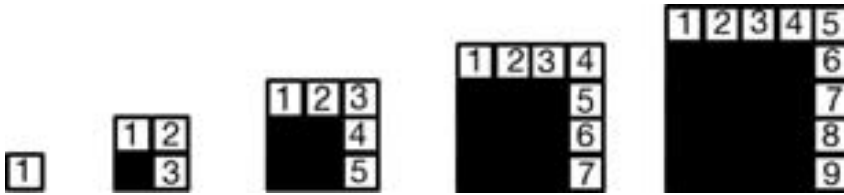
8.6. gyakorlat

A fenti regiszteres gépi program kiszámítja a négyzetes függvényt. Milyen algoritmust valósít meg?

Kiderül, hogy az n -edik négyzetszámot úgy találjuk meg, ha egyszerűen az első n páratlan számot összeszámoljuk. Tehát 1^2 az 1, $2^2 = (1 + 3) = 4$, $3^2 = (1 + 3 + 5) = 9$ stb. A 8.2. ábra megvilágítja, hogy miért van ez így. A fenti négyzetelésre szolgáló regiszteres gépi program egy algoritmus, amely a páratlan számok sorozatának összegét veszi az n -edik kifejezésig (ahol n a függvény bemenete - a szám, amellyel az R-ben kezdünk₁).

8.7. gyakorlat

Tervezz egy olyan regiszteres gépi programot, amely kiszámítja az $f(x) = 2x^2 - 1$ függvényt (Ennek a megoldása rajtad múlik.)



8.2. ábra A négyzetek progressiója.

Mielőtt továbblépnénk, érdemes néhány érdekességet kiemelni az ebben a fejezetben eddig tárgyalt anyagból.

Most olyan regiszteres gépi programokat néztünk meg, amelyek két különböző módon képesek másolni a regiszterek tartalmát, összeadást és szorzást, valamint négyzetszámítást végezni. A négyzetszámítási függvény kiszámításának utóbbi módja egy sorozat tagjainak generálását és felhalmozását jelentette.

Mindezt az alapvető erőforrások - inkrementáló és dekrementáló utasítások - igen gyér készletéből építették fel. Láthatjuk, hogy ezeket hogyan lehet tovább építeni meglehetősen összetettebb számításokba - olyan programokba, amelyek nagyon összetett algoritmusokat valósítanak meg. Erre a fogalomra a 10. fejezetben térünk vissza.

Egy másik érdekes pont annak bemutatása, hogy egy függvény kiszámításának több módja is lehet. Valójában a nagyon bonyolult függvények valószínűleg számos algoritmussal kiszámíthatók. Ezt a másik fogalmat a 10. fejezetben fogjuk újra megvizsgálni.

A következő fejezetben megnézzük, hogyan használhatjuk a Gödel-kódolást a programokon belüli hivatkozások megkönnyítésére. Ezzel a módszerrel pontosan látni fogjuk, hogy mennyi mindent lehet elérni csak inkrementáló és dekrementáló utasításokkal.

Először azonban itt van egy utolsó kihívás azok számára, akiknek felkeltettük az érdeklődését.

8.8. gyakorlat (Kihívás)

Láttuk, hogyan lehet egy egyszerű sorozat tagjait felhalmozni. A tagok felhalmozása helyett könnyen módosíthatjuk a programot úgy, hogy az n -edik tagot adja vissza (ahol n a függvény bemenete), más szóval, hogy az R_1 az első n páratlan szám összege helyett az n -edik páratlan számmal fejeződjön be.

A Fibonacci-sorozat a következőképpen

alakul: 0, 1, 1, 2, 3, 5, 8, 13, 21, 34, 55, 89, .

..

Legyen $\text{fib}(x)$ az a függvény, amelynek kimenete az x -edik Fibonacci-szám.

Írjon egy regiszteres gépi programot a $\text{fib}(x)$ kiszámítására.

Hint: $\text{fib}(1) = 0; \text{fib}(2) = 1; \text{fib}(x + 1) = \text{fib}(x-1) + \text{fib}(x)$

9. FEJEZET

UNIVERZÁLIS GÉPEK

Ez a fejezet befejezi a számításelmélet ismertetését. Eddig a formális rendszereket széles körben tárgyaltuk, és a formális rendszerek egy speciális fajtáját - a *regisztergépet* - használtuk a *számítás* és a *kiszámíthatóság* meghatározására.

Ennek során némi betekintést nyertünk abba, hogy mi az, ami kiszámítható, és mi az, ami nem. Láttuk például a 8.4. szakaszban, hogy nincs effektív eljárás annak meghatározására, hogy egy adott program meg fog-e állni vagy sem. Mivel a regisztergépeket effektivitási korlátok kötik, azt mondhatjuk, hogy ezért nincs olyan regisztergép, amely bármely adott programról meg tudja állapítani, hogy az meg fog-e állni vagy sem. Következésképpen, mivel a kiszámíthatóságot a regisztergépek szempontjából definiáljuk, azt mondhatjuk, hogy a *megállási probléma* nem kiszámítható.

A 8.5. szakaszból azt is tudjuk, hogy a számítás tárgyai *függvények*, és pontos formális definícióval rendelkezünk arra vonatkozóan, hogy mit jelent egy függvény *kiszámítása*. Tehát tudjuk, hogy legalább *néhány* függvény kiszámítható, és hogy *csak* a függvények kiszámíthatók.

Még mindig vannak azonban fontos kérdések, amelyek megválaszolatlanul maradtak. Egyrészt nem tudunk számot adni a számítás *korlátairól*. Még nem tudjuk pontosan, *hogy mely* függvények számíthatóak.

Az *effektivitás* és az *algoritmizálhatóság* formális leírása is hiányzik. Emlékezzünk vissza a 7.2. szakaszból, hogy az effektivitást csak *informálisan* definiáltuk az algoritmizálhatóság szempontjából. Azt mondtuk, hogy egy eljárás csak abban az esetben effektív, ha van algoritmus a végrehajtásához. Az algoritmus fogalmát azonban szintén csak informálisan konkretizáltuk, egy mechanikus eljárás vagy utasításkészlet értelmében, amelyet mindenféle megértés vagy értelmezés nélkül végre lehet hajtani.

Ebben a fejezetben először is pontos választ adunk a számítás határainak kérdésére azáltal, hogy az effektivitás *informális* fogalmát összekötjük a kiszámíthatóság *formális* leírásával, és

ezáltal megadjuk az algoritmizálhatóság formális definícióját.

Ezzel a beszámolóval a kezünkben a továbbiakban egy olyan gép ~~át~~ fogjuk kidolgozni, amely képes *bármilyen* kiszámítható függvényt - egy *számítógépet* - kiszámítani.

9.1 CHURCH/TURING TÉZIS

A *megállási probléma a döntési probléma* - vagy *döntési probléma* - egy sajátos esete, amely már jóval a számítás formális elmélete előtt is érdekelte a ~~matematikus~~ és a logikusokat. A döntési probléma egy adott formális rendszerre vonatkozóan arra a kérdésre utal, hogy létezik-e effektív eljárás annak meghatározására, hogy a rendszer bármely adott állapotáról megállapítható-e, hogy az a rendszerben *keletkezik-e* vagy sem. Ha van ilyen eljárás, akkor a rendszert *eldönthetőnek* mondjuk.

Csak azon olvasók kedvéért, akik esetleg értenek a modern logikához, hadd mondjam el a következőket. Az a kérdés, hogy egy adott logikai rendszer *eldönthető-e* vagy sem, az a kérdés, hogy van-e módszer arra, hogy a rendszer nyelvének bármely tetszőleges mondatára megmutassuk, hogy az a mondat bizonyítható-e a rendszerben vagy sem. A *bizonyíthatóság* fogalma azonban a *generálás* fogalmára redukálódik. Ha a rendszer kezdeti állapotát a rendszer axiómáinak, a rendszer bizonyítási elméletét pedig a szabályainak tekintjük, akkor a *eldönthetőség* egyszerűen azt jelenti, hogy bármely tetszőleges ~~mondat~~ meg lehet-e mutatni, hogy az szerepel-e a kezdeti állapotból a szabályoknak megfelelően levezetett valamelyik állapotban. Azok az olvasók, akik elvégeztek egy első logikai kurzust, tudják, hogy az állítmányi kalkulus *eldönthető*, de a *predikátumkalkulus* nem.

Ne aggódjon, ha az előző bekezdés keveset vagy semmit sem jelent az Ön számára. Csupán mellékszálnak szánom azoknak, akiknek van némi érzékük a logikához.

Alan Turing brit matematikus 1936-ban megjelent, *On Computable Numbers, with an Application to the Entscheidungsproblem* (A *kiszámítható számokról, egy alkalmazással a döntési problémára*) című tanulmányában azzal érvelt, hogy egy bizonyos logikai rendszer nem *eldönthető* (a *predikátumszámítás* azok számára, akiknek ez jelent valamit), és ezzel megalapozta a mai számítástechnikát.

Turing a *kiszámíthatóságot* a ma *Turing-gépek* néven ismert elméleti gépek segítségével határozta meg. A számításelméleti irodalomban sokkal gyakrabban beszélnek Turing-gépekről, mint regisztergépekről. A Turing-gépek *kiszámíthatósága* és a regisztergépek *kiszámíthatósága* azonban bizonyíthatóan egyenértékűek - minden és csakis ugyanazokat a függvényeket képesek kiszámítani.

Bárhol, ahol Turing-gépekről vagy Turing-számíthatóságról beszélnek, ezt felcserélhetjük a regisztergépekkel és a regisztergépek számíthatóságával.

Két okból döntöttem úgy, hogy a számításelmélet kezdeti bemutatásához regiszteres gépeket használok. Először is, az egyszerű regisztrálógépek könnyen megvalósíthatók fizikai eszközökkel, érmékkel (érmék, gombok stb.). Ez nagy segítséget nyújthat a működésük megértéséhez. Másodsor, bár a regisztergépek és a Turing-gépek funkcionálisan egyenértékűek, a regisztergépek műveleteinek mechanikája közelebb áll ahhoz, ahogyan a számítást ténylegesen megvalósítják az általunk ismert számítógépekben.

1936-os tanulmányában Turing azt állította, hogy minden, és csakis azok az eljárások, amelyek algoritmikusak, kiszámíthatók Turing-(regiszter)gépekkel.

Egy másik, 1936-ban megjelent tanulmányában egy amerikai matematikus, Alonzo Church azt állította, hogy az effektív számítási képesség informális fogalma (legalábbis a pozitív egész számítások esetében) a *rekurzív függvény* formális fogalmával érthető meg. Mindkét esetben Church és Turing célja az volt, hogy egy informálisan jól ismert fogalmat - az algoritmizálhatóság vagy effektivitás fogalmát - egy olyan fogalomhoz kösse, amelyre pontos formális definíció létezik.

A Church/Turing-tézis, ahogyan ma már általánosan hivatkoznak rá, röviden így fejezhető ki: *minden és csakis effektív eljárás kiszámítható függvény.*

Fontos megérteni, hogy a "kiszámítható" kifejezés használata a Church/Turing tézisben a kiszámíthatóság egy bizonyos formális értelmezésére utal, nevezetesen a *nyilván tartógépek kiszámíthatóságára*, ahogyan azt mi definiáltuk. Ez talán nem tűnik túl fontosnak, de a Church/Turing-tézis félreértelmezésének jelentős lehetősége van, ha ezt nem tartjuk szem előtt, és a szakirodalomban gyakran találkozunk ilyen félreértésekkel.

Azt is érdemes kiemelni, hogy az effektív eljárások és a kiszámítható függvények közötti ekvivalencia *tézis* és nem *tétel*. Más szóval, ez egy javasolt ekvivalencia, de nem *bizonyított*. Úgy tűnik azonban, hogy a tétel természetéből adódóan kevés esély van a tényleges bizonyítására, mivel egy formális fogalmat egy informális fogalommal tesz egyenlővé. A tétel egyik iránya - miszerint minden kiszámítható függvény effektív eljárás - nyilvánvaló, mivel a kiszámíthatóságot effektív eljárásokkal definiáltuk. A másik irány - hogy csak a kiszámítható függvények effektív eljárások, vagy hogy minden effektív eljárás kiszámítható függvény - kevésbé nyilvánvaló. Az azonban tény, hogy 1936 óta jelentős mennyiségű

a tézis mellett szóló bizonyítékokat, és nem talákoztam ellenpéldákkal.

A Church/Turing tézis tehát választ ad a fejezet elején feltett két kérdésre. Most már pontosan tudjuk, hogy mely függvények számíthatóak - mind és csakis azok, amelyek algoritmikusak. Most már rendelkezünk az effektivitás/algorithmizálhatóság *formális* jellemzésével is - pontosan azzal, amit a 8.5. szakaszban a komputabilitásra adtunk.

A Church/Turing tézis szerint minden adott algoritmushoz létezik egy regisztergépes program (mivel minden effektív eljárás regisztergépen kiszámítható). Ha ehhez hozzávesszük, hogy a regisztergépprogramok determinisztikus formális rendszerek, és hogy bármely adott determinisztikus formális rendszer műveletei algoritmikusak, akkor láthatjuk, hogy az *algoritmikus eljárások, effektív eljárások, regisztergépprogramok, determinisztikus formális rendszerek és kiszámítható függvények* mind egyenértékű beszédmódok - mind ugyanazt az osztályt specifikálják.

9.2 GÖDEL KÓDOLÁS

Láttuk, hogy bármilyen algoritmushoz, amelyet meg tudunk határozni, megtervezhetünk egy regiszteres gépi programot a kiszámításához. E fejezet végső célja, hogy megadjunk egy olyan regisztergépprogramot, amely *maga is* képes bármely algoritmus kiszámítására. Más szóval, olyan regisztergépprogramot akarunk megadni, amely képes bármely más regisztergépprogramot *emulálni* vagy *példázni*.

Ez azt jelenti, hogy valamilyen módon alkalmassá kell tennünk a regisztergép programjait arra, hogy egy másik regisztergép bemenetét képezzék. Mivel a regisztergépek bemenetei természetes számok, szükségünk lesz valamilyen módszerre, hogy a regisztergépek programjait természetes számokként *kódoljuk*. Továbbá, ennek a kódolási eljárásnak egyedinek és hatékonynak kell lennie. Más szóval, amellet, hogy a kódolási eljárásnak algoritmikusnak kell lennie, olyannak is kell lennie, hogy bármely adott regisztergépprogram egyedi kódot eredményezzen, és fordítva. Az általunk használt eljárást *Gödel-kódolásnak* nevezzük.

Kurt Gödel zseniális logikus és matematikus volt, aki a huszadik században a matematikai logika legfontosabb eredményeit adta. Gödel 1931-ben közzétett befejezetlenségi tételei vitathatatlanul azok a tételek, amelyeket a nem matematikusok és logikusok is a legjobban ismernek, ugyanakkor a legkevésbé értik őket. Valójában a Gödel-eredményekről meglehetősen sok ostobaságot írnak az elmefilozófiában.

Sajnos a Gödel-féle befejezetlenségi tételek megfelelő megértéséhez, hogy lássuk, hogyan értelmezik őket gyakran félre a komputualizmus elleni érvekben, jóval nagyobb matematikai kifinomultságra lenne szükség, mint amit a bevezető olvasótól, akinek ez a kötet szól, elvárnánk. Suffice it to say, ha a komputationalizmus elleni érvekkel találkozunk, amelyek ezeket a tételeket használják, akkor ezeket a tételeket nem kell félreértetni.

A gödéli kódolás módszere egyike azoknak a nagyon ügyes trükköknek, amelyeket Gödel a befejezetlenségi tételek bizonyítására használt. Megkönnyíti a formális rendszer elemeire való hivatkozást a rendszeren belülről, ami pontosan az, amire szükségünk van. Szerencsére a mi céljaink szempontjából a Gödel-féle kódolás módszerének megértése nem igényel nagyobb matematikai kifinomultságot, mint a szorzás és a szorzás, valamint a prímszám fogalmának ismerete.

A négyzetelés és a kockaegyengetés matematikai műveletei példái a *hatványozásnak*, amely egy számnak egy másik szám *hatványára* vagy *exponensére való* emelésének általánosabb művelete. Az a^b kifejezést úgy értékeljük ki, hogy a -t b -szer megszorozzuk önmagával - a az *alap*, b pedig az *exponens*.

A *prímszám* olyan szám, amelynek nincs más tényezője, mint önmaga és 1. A prímszámok érdekes - és a mi céljaink szempontjából döntő - tulajdonsága, hogy minden természetes szám egyértelműen kifejezhető prímszámok szorzataként. Ennek bizonyítása kissé bonyolult, ezért itt egyszerűen csak a szavamat kell adnunk. Továbbá a prímtenyezőket algoritmikusan lehet kiszámítani (számos ismert algoritmus létezik). A részletek ismét bonyolultak, így itt is csak a szavamat kell adnod.

Tehát bármely adott természetes szám esetén effektíven és egyértelműen meghatározhatjuk prímtenyezőit. A Gödel-kódolás módszere ezt használja ki azáltal, hogy a programokat prímszámok exponenseként kódolja.

Emlékezzünk vissza, hogy egy regiszteres gépi program kétféle formájú utasítássorozat - $I a b$ vagy $D a b c$. A növelő utasításokhoz három információt kell kódolnunk - hogy ez egy növelő utasítás, valamint az a és b számokat. A csökkentő utasításokhoz négy információt kell kódolnunk - hogy ez egy csökkentő utasítás, valamint az a , b és c számokat. A 2-vel kezdjük, ami azt jelzi, hogy a kód egy inkrementáló utasítás kódja. Ezt megszorozzuk az a és a 7 hatványára emelt 5-tel.

b hatványára emelve.

A dekrementáló utasítások kódolásához négy prímszámot használunk. A 3-mal kezdjük, ami azt jelzi, hogy a kód egy dekrementáló utasítás kódja.

utasítás. Ezt megszorozzuk az a hatványára emelt 5-tel, a b hatványára emelt 7-tel és a c hatványára emelt 11-tel.

Hogy ezt szimbolikusan kifejezzük:

$I a b$ kódja $2 \cdot 5a \cdot 7b$ (ahol a jelöli a szorzás) és $D a b c$

kódja $3 \cdot 5a \cdot 7b \cdot 11c$

(Ezeket a kifejezéseket mindig exponenciális formában hagyjuk.)

Vegyük észre, hogy a növelő utasítás kódja mindig páros szám, a csökkentő utasítás kódja mindig 3-mal osztható, és minden olyan szám, amely nem osztható 2-vel vagy 3-mal, nem képezi egyetlen utasítás kódját sem.

Ha $\langle I$ valamilyen utasítást jelöl, akkor a $\# \langle I$ -t használjuk a $\langle I$ kódjára való hivatkozásra. Tehát, hogy néhány példát mondjak:

$ha \langle I = I 2 \quad 3then \# \langle I = 2 \cdot 5^2 \cdot 7^3$
 $ha \langle I = I 7 \quad 4den \# \langle I = 2 \cdot 5^7 \cdot 7^4$
 $if \langle I = D 3 8 5 then \# \langle I = 3 \cdot 5^3 \cdot 7^8 \cdot 11^5$
 $if \langle I = D 2 6 3 then \# \langle I = 3 \cdot 5^2 \cdot 7^6 \cdot 11^3$

és így tovább.

9.1. gyakorlat

Kódolja a nyolcadik fejezet feladatmegoldásainak utasításait. (A sorszámokat egyelőre hagyjuk figyelmen kívül - csak az utasításokat kódoljuk.)

Most, hogy már tudjuk, hogyan kell utasításokat kódolni, pontosan ugyanezt az eljárást fogjuk használni egész programok kódolására.

A program egy sorszámból és egy utasításból álló n sorszámú program. Az egyszerűség kedvéért feltételezzük, hogy a sorok számozása a természetes számok 1-től kezdődő sorozatában történik. Amikor a programkódokat az őket alkotó utasításkódokra dekódoljuk, egyszerűen ugyanebben a sorrendben soroljuk fel őket. Ez lehetővé teszi, hogy a kódokban figyelmen kívül hagyjuk a sorszámokat.

Adott egy n utasításból álló sorozat, ezt úgy kódolhatjuk, hogy egyszerűen többszörözzük az első n prímszámot, amelyek mindegyikét felemeljük az adott utasítás kódját jelentő exponensre. Tehát az első prímszámot az első utasítás kódjának megfelelő exponensre emeljük, a második prímszámot a második utasítás kódjának megfelelő exponensre emeljük, és így tovább.

Hogy ezt szimbolikusan kifejezzük:

adott egy program $1 < I_1, 2 < I_2, 3 < I_3, \dots, n < I_n$
 kódolható $2^{\# < I_1} \cdot 3^{\# < I_2} \cdot 5^{\# < I_3} \cdot \dots \cdot p^{\# < I_n}$ (ahol p a
 n -edik prímszám).

Ez meglehetősen bonyolultnak tűnhet, de néhány példa megmutatja, hogy valójában nagyon egyszerű. A kifejezéseket mindig egyszerűen exponenciális formában hagyjuk, és a dolgok könnyebb olvashatósága érdekében változókat rendelünk az utasításkódokhoz, ahogy az alábbi példák is mutatják.

A program:

```
1 D 3 1 2
2 D 1 3 5
3 D 2 4 5
4   I 32
```

kódolva:

$2^a \cdot 3^b \cdot 5^c \cdot$

7^d ahol :

$$a = 3 \cdot 5^3 \cdot 7^1 \cdot 11^2$$

$$b = 3 \cdot 5^1 \cdot 7^3 \cdot 11^5$$

$$c = 3 \cdot 5^2 \cdot 7^4 \cdot 11^5$$

$$d = 2 \cdot 5^3 \cdot 7^2$$

A program:

```
1 D 3 1 2
2 D 4 2 3
3 D 1 4 9
4 D 2 5 7
5 I 3 6
6 I 4 4
7 D 4 8 3
8 I 2 7
9 D 3 10 12
10 I 1 9
```

kódolva:

$2^a \cdot 3^b \cdot 5^c \cdot 7^d \cdot 11^e \cdot 13^f \cdot 17^g \cdot 19^h \cdot 23^i \cdot 29^j$

ahol:

$$a = 3 \cdot 5^3 \cdot 7^1 \cdot 11^2$$

$$b = 3 \cdot 5^4 \cdot 7^2 \cdot 11^3$$

$$c = 3 \cdot 5^1 \cdot 7^4 \cdot 11^9$$

$$d = 3 \cdot 52 \cdot 75 \cdot 117$$

$$e = 2 \cdot 53 \cdot 76$$

$$f = 2 \cdot 54 \cdot 74$$

$$g = 3 \cdot 54 \cdot 78 \cdot 113$$

$$h = 2 \cdot 52 \cdot 77$$

$$i = 3 \cdot 53 \cdot 710 \cdot 1112$$

$$j = 2 \cdot 51 \cdot 79$$

9.2. gyakorlat

A fenti második példa a 8.4. feladat megoldását kódolja.
Kódolja be a 8. fejezet többi feladatának megoldását.

Bár a programkódokat exponenciális formában hagyjuk, azok természetes számokra oldódnak fel. Igaz, hogy felbontásuk nagyon nagy természetes számok, de ettől függetlenül természetes számok. Mint ilyen, most már rendelkezünk egy egyedi és effektív eljárással, amellyel bármilyen hosszúságú regiszteres gépi programot egyetlen természetes számként kódolhatunk, amely egyedileg és effektíven dekódolható. Ez azt jelenti, hogy most már van egy mechanizmusunk a programokon belüli programokra való hivatkozásra, mivel egy reg- ister tárolhatja egy program kódját.

9.3. gyakorlat

Adjon három példát olyan számokra, amelyek differens okokból nem programkódok.

9.4. gyakorlat (Kihívás)

Mi a legkisebb olyan természetes szám, amelyik programkód?

9.3 EGY UNIVERZÁLIS GÉP

A Gödel-kódolás módszerével felvértezve, most már készen állunk arra, hogy leírjuk azt a regiszteres gépi programot, amely bármilyen algoritmikus függvényt képes kiszámítani.

Legyen [UM] az a regisztergép, amelynek programját a következő eljárás írja le.

Először dekódoljuk az R_1 tartalmát. Ha az R_1 nem tartalmazza egy program kódját, akkor álljunk meg. Ha az R_1 tartalmazza valamilyen P program kódját, akkor adjunk hozzá 1-et a P -ben említett összes regiszterhez, majd futtassuk le a P -t, és a program befejezésekor másoljuk az R_2 tartalmát az R_{1-be} .

Nyilvánvaló, hogy ez az eljárás effektív, ezért a Church/Turing-tézis szerint létezik ilyen regiszteres gépi program.

Ha a [UM] programot a $\#P, a_1, \dots, a_n$ az első n + 1 regisztert, ugyanolyan eredményt adna, mintha egy gépet futtatnánk a P programmal és az a bemenettel a_1, \dots, a_n első n regiszterben.

Ezért az [UM] eljárás során az összes értékhez hozzáadunk egyet. regiszterek P-ben (ha $R_1 \#P$ -re dekódol), és miért másoljuk $R_2 \leftarrow R_1$ -be P ter- minációjakor - mert P bemenetei mind egy regiszter mentén tolnának el [UM].

Következésképpen az [UM] bármilyen regiszteres gépi programot emulálhat. A program kódját egyszerűen az első regiszterbe, a függvény n bemenetét pedig a következő n regiszterbe helyezzük, és a [UM] programot a leírtak szerint futtatjuk.

A regisztergépet [UM] *univerzális gépnek* nevezik - bármilyen regisztergéppel kiszámítható függvényt képes kiszámítani. Egyszerűbben fogalmazva, az [UM] egy *számítógép*.

A modern digitális számítógépek, ahogyan mi ismerjük őket, instantiált uni- versális gépek. A személyi számítógépek, a nagyszámítógépek és még a szuperszámítógépek sem nagyobb teljesítményűek, mint a [UM] - nincs semmi, amit a [UM] ne tudna kiszámítani. Valójában az [UM] nagyobb teljesítményű, mint bármely fizikai gép, mivel ez egy elméleti idealizáció, amelyet nem korlátoz a fizikai megnyilvánulás. Erről a következő fejezetben még többet fogunk mondani.

A számításelméleti áttekintésünket most befejeztük. A számítógépek mibenlétéről alkotott új ismereteinkkel felvértezve ideje visszatérnünk a filozófiai anyaghoz, és megvitatnunk azt az elmeelméletet, amelynek értékelése központi feladatunk - a komputacionalizmust.

10. FEJEZET

COMPUTATIONALISM

Eddig azt a kérdést vizsgáltuk, hogy mi lehet az elme, és ennek során számos olyan filozófiai elméletet vizsgáltunk meg, amelyek arra törekedtek, hogy választ adjanak erre a kérdésre. A funkcionalizmus bevezetése után felfüggesztettük ezt a ~~vá~~ és effortáinkat arra fordítottuk, hogy kidolgozzunk egy szigorú beszámolót arról, hogy mi a számítás. Itt az ideje, hogy ott folytassuk, ahol abbahagytuk.

Ebben a fejezetben a komputációs elméletnek az előző fejezetben felépített megértését fogjuk felhasználni a funkcionalista keretrendszer sajátos módon történő kibontásához.

Többféleképpen lehetünk funkcionalisták, attól függően, hogyan elemezzük a mentális állapotok funkcióit. Célunk az, hogy tisztességes és pontos feldolgozást adjunk annak az elméletnek, hogy ezeket a funkciókat *számítási* szempontból kell kiteljesíteni.

Ennek során nyilvánvalóvá válik a komputációs elmélet kidolgozásának haszna. Először is, most már beszélhetünk a számításról anélkül, hogy egyszerűen csak laza beszédbe bocsátkoznánk - pontos formális definíció és néhány finom megkülönböztetés áll rendelkezésünkre. Ez lehetővé teszi számunkra, hogy lássuk, hogy bizonyos, a komputualizmus ellen néha felhozott ellenvetések valójában nem az elméletet célozzák - *szalmabábukat* célozzák, a *számítógépekről* alkotott, nem eléggé kifinomult felfogás alapján.

A *szalmabábu* célba vétele - vagy a *szalmabábu tévedés* elkövetése - azt jelenti, hogy egy ellenérvvel szemben álló álláspontot gyengébbnek minősítünk, mint amilyen valójában, majd a gyengébb álláspont ellen érvelünk. Egy gyengébb tévhit ellen érvelni egyáltalán nem használ a tényleges ellenérvekkel szemben. Ezért, ha valaki ezt a tévedést elköveti, akkor azt mondjuk, hogy szalmabábút épít, pusztán azért, hogy ledöntse azt.

A komputualizmus egyértelmű jellemzésével kezdjük. Ezt követően elsődleges célunk, hogy foglalkozzunk az elmélettel kapcsolatos néhány lehetséges tévhittel, és tisztázzuk azokat. Ezt

követően tárgyaljuk

pontosan mi mellett kötelezi el magát egy számítástudós, és fontoljunk meg néhány közvetlen következményt.

Ebben a fejezetben a komputacionalizmus néhány érdemét is megvitatjuk, és foglalkozunk az elmélettel szembeni néhány *prima facie* ellenvetéssel. A könyv hátralévő részében a komputacionalizmus teljesebb értékelésével foglalkozunk majd. Megnézzük, hogy a kognitív tudományágakból származó eltérő anyagok hogyan befolyásolják jelentősen az elmélet tarthatóságát.

Foglalkozunk most azzal a kérdéssel, hogy pontosan mit takar az az állítás, hogy a mentális állapotok számítási állapotok.

10.1 AMI A KOMPUTACIONALIZMUS NEM

A komputacionalizmus az a nézet, amely szerint az elme létezése egy bizonyos formális rendszer vagy rendszerek gyűjteménye. Más szóval, mivel a mentális műveleteket formális rendszerek műveleteinek tekintik, a mentális műveleteket *számításoknak* tekintik. Tehát az elmével rendelkezni, állítja a komputacionalista, *csak annyit jelent*, hogy bizonyos számítási folyamatokban veszünk részt.

A komputacionalizmus egyértelműen a funkcionalizmus egy fajtája. A funkcionalista szerint az állapotok kizárólag a bemenetek, kimenetek és más mentális állapotok közötti kapcsolatok közvetítésében betöltött jellegzetes funkcióik alapján mentálisak. A komputacionalizmus egyszerűen csak egy módja annak, hogy ezeket a közvetítő kapcsolatokat konkretizáljuk - a kérdéses kapcsolatokat számításoknak tekintjük.

A komputacionalizmus *nem az a* nézet, hogy a formális rendszerek műveletei önmagukban mentális műveletek. Vagyis nem az a nézet, hogy az elme meglétéhez egyáltalán szükséges bármilyen formális rendszer instanciája. Ez nyilvánvalóan túlságosan lekötelezi a komputacionalistákat, mivel ez azt követelné meg tőlük, hogy mindenféle tárgyaknak - termosztátoknak, traffic lámpáknak, kézi elektronikus játékoknak - nyilvánvalóan nevetséges módon mentalitást tulajdonítsanak.

A komputacionalistákkal szemben tisztázzuk tehát, hogy ők csak azt a nézetet vallják, hogy egy *bizonyos* formális rendszer - nevezzük azt [ÉSZ] - példányosítása szükséges az elme meglétéhez.

Ez inkább a komputacionalizmus erős megfogalmazása. Egy komputacionalista azt vallhatja, hogy nem egyetlen átfogó formális rendszert kell azonosítani, hanem inkább azt, hogy a mentalitás bizonyos számú különböző algoritmus függvénye. Én a mentalitás egy sajátos - általam meglehetősen intuitívnek vélt - megértését szeretném előmozdítani, amely egyetlen átfogó formális rendszert feltételez, ezért további értesítésig ezzel az erős tézissel fogok

dolgozni. Ne feledje azonban, hogy a

Az általam leírt elmélet nem az egyetlen történet, amely egy komputationalista számára elérhető. Ha azonban az elmélet egy erős változata védhetőnek bizonyul, akkor a fortiori bármely gyengébb változata is védhető.

A komputationalista tehát nem elkötelezett amellett a nézet mellett, hogy a személyi számítógép műveletei mentális műveletek. Nem elkötelezett amellett a nézet mellett sem, hogy a nagyon nagy teljesítményű számítási eszközöknek, például a szuperszámítógépeknek van elméjük. Ő csak azt vallja, hogy ha valamilyen szubsztrátum futtatja a programot [MIND], akkor ez a szubsztrátum ezáltal rendelkezik elmével.

Most három különböző módon hivatkoztam a [MIND]-re: formális rendszerként, algoritmusként és regiszteres gépi programként. Emlékezzünk vissza a 9. fejezetből, hogy a Church/Turing-tézis szerint ez a három egyenértékű beszédmód. Ha a [MIND] egy algoritmus, akkor azt egy regisztergépprogrammal lehet megvalósítani (ami éppen egy determinisztikus formális rendszer).

Tisztázzunk egy másik lehetséges tévhitet az elmélettel kapcsolatban. A komputationalista állítás egy másik fontos módon is eltér attól a nézettől, hogy a személyi számítógépeknek lehet vagy van elméjük. A modern digitális számítógépek, mint az előző fejezetben láttuk, instantiált *univerzális gépek*. A komputationalista *nem azt* állítja, hogy az elmével rendelkezni annyi, mint egy univerzális gépet instanciálni. Azt állítja, hogy az elmével rendelkezni annyit jelent, mint egy bizonyos *regiszteres gépet* - nevezetesen a [MIND]-et - instanciálni.

Itt néhány ponton óvatosságnak kell lennünk. Először is, meg kell értenünk, hogy bár a digitális számítógépek, ahogyan mi ismerjük őket, instanciált univerzális gépek, de *tökéletlenül* instanciáltak. Az univerzális gépek olyan elméleti eszközök, amelyek erőforrásai - bár a feltételezés szerint végesek - egyébként korlátlanok. Az instanciált univerzális gépek fizikai eszközök, amelyeket fizikai korlátok kötnek. Míg tehát az univerzális gépek *elvileg* bármilyen programot képesek futtatni (a fortiori képesek futtatni a [MIND]-et), addig az instanciált univerzális gépeket a *gyakorlatban* fizikai korlátaik korlátozzák, és mint ilyenek, előfordulhat, hogy hardveres szinten nem rendelkeznek megfelelő számítási erőforrásokkal bizonyos programok, például a [MIND] futtatásához.

Tehát bizonyos értelemben nem teljesen helyes azt mondani, hogy a modern digitális számítógépek instanciált univerzális gépek, mivel lehetnek programok, amelyek túlmutatnak a számítási erőforrásaikon. A digitális számítógépek, ahogyan mi ismerjük őket, az univerzális gépek *közelítései*. A számítási hardverek

egymást követő generációi egyre közelebbi és közelebbi közelítéseket nyújtanak, mivel egyre nagyobb és nagyobb számítási erőforrásokat biztosítanak.

Még egy univerzális gép tökéletes megvalósítása egy szubsztrátumban sem lenne önmagában *elégéses* ahhoz, hogy a szubsztrátumnak elméje legyen. A

A szubsztrátnak ekkor a megfelelő programot kell futtatnia, mivel az elmével rendelkezni - mondja a komputationalista - annyit jelent, hogy a program [LELK] *működésben van*.

A *gyakorlati számítási teljesítményt* két paraméter - a tárhely és a feldolgozási sebesség - mentén tudjuk megadni. Egy adott program igényei akkor mondhatjuk, hogy meghaladják egy adott eszköz gyakorlati ~~számítási~~ teljesítményét, ha az igényei meghaladják az eszköz tárolókapacitását vagy számítási sebességét (vagy mindkettőt).

Tehát míg a komputationalista elkötelezett amellett, hogy bármely univerzális gép (korlátozások nélkül) képes futtatni a [MIND]-et - és ezáltal elmével rendelkezne -, addig *ipso facto* nem elkötelezett amellett, hogy bármely létező digitális számítógépnek lehet elméje. Egy komputacionalista úgy gondolhatja, hogy a [MIND] követelményei meghaladják a jelenleg rendelkezésre álló (nem biológiai) számítási eszközök (néhányikének vagy mindegyikének) gyakorlati számítási teljesítményét.

Következésképpen egy komputationalistának nem kell azt gondolnia, hogy a személyi számítógépnek lehet elméje. Minden valószínűség szerint azt fogja állítani, hogy fizikai korlátai miatt nem lehet. A komputeralista azt a nézetet vallja, hogy az emberi agy biztosítja a biológiai komputációs hardvert az [ÉSZ] emberi megvalósításához. Tekintettel arra, amit az emberi agy rendkívüli tárolókapacitásáról és működési sebességéről tudunk, a komputeralista valószínűleg azzal érvel, hogy bármilyen nem biológiai számítási eszköznek, amely elég erős ahhoz, hogy az [ÉSZ] működtesse, (legalábbis megközelítőleg) az emberi agy tárolókapacitásával és működési sebességével kell rendelkeznie. A digitális számítógépek, bármennyire is egyre nagyobb teljesítményűek, még mindig nem érik el a közelébe.

Mielőtt továbblépünk, még egy utolsó lehetséges félreértést kell tisztáznunk. Emlékezzünk vissza a 7.1. szakaszból, hogy ahhoz, hogy egy eljárás hatékony legyen, elvileg képesnek kell lennie arra, hogy egy ember - megfelelő idővel - elvégezze, ha csak egy halom követ (vagy papírt és ceruzát) használ, és a feladathoz nem hozza magával a megértést. Az univerzális gépek műveletei teljesen effektívek, így az emberi elme többek között úgy közelítheti meg az univerzális gépet (bár jelentős korlátozásokkal), hogy aktívan, lépésről lépésre végigmegy bármelyik adott regiszteres gépprogram műveletein. Következésképpen az egyik dolog, amire egy [MIND]-et futtató eszköznek képesnek kell lennie, hogy megközelítsen egy univerzális gépet (legalábbis olyan meglehetősen gyenge módon, ahogyan az ember képes rá).

Ez azonban nem jelenti azt, hogy egy univerzális géphez való

közelítés *elegendő* az elme meglétéhez. Éppen ellenkezőleg. Ez azt jelenti, hogy a komputualizmus státuszától függetlenül az elme megléte alkalmas arra, hogy (nagyon gyengén) megközelítsünk egy univerzális gépet.

Foglaljuk össze az eddig tárgyalt lehetséges zavaró pontokat. Először is, egy komputationalista nem kötelezi el magát azon nézet mellett, hogy minden számítás mentális művelet. Ők azon nézet mellett kötelezik el magukat, hogy *bizonyos* számítások - azok, amelyek az [ÉN] műveletei - mentális műveletek.

Másodszor, egy komputationalista nem elkötelezett amellett a nézet mellett, hogy egy univerzális gép példányosítása szükséges az elme meglétéhez. Ők azt a nézetet vallják, hogy egy tökéletesen instanciált (korlátozások nélküli) univerzális gép *képes* arra, hogy elmével rendelkezzen. Mivel az elme többek között képes egy univerzális gépet közelíteni, a komputationalista elkötelezett amellett is, hogy a [MIND]-t futtató bármely számítási eszköz képes egy univerzális gép gyenge közelítésére.

Végezetül, egy komputationalista nem kötelezi el magát amellett a nézet mellett, hogy bármely adott számítási eszköz képes a [MIND] megvalósítására, mivel a programnak olyan követelményei lehetnek, amelyek meghaladják az adott eszköz gyakorlati számítási erőforrásait. Következésképpen egy komputationalista boldogan tagadhatja, hogy egy olyan eszköznek, mint az Ön személyi számítógépe - amely egy univerzális géphez való közelítés - valaha is lehetne elméje. A komputationalista azonban elkötelezett amellett, hogy azt állítsa, hogy bármely fizikai eszköz, amely elegendő gyakorlati számítási kapacitással rendelkezik a [MIND] futtatásához, képes elmével rendelkezni. Az, hogy pontosan milyen számítási erőforrásokra van szüksége az [ÉSZ] számára, empirikus felfedezés kérdése. Most, hogy gondosan azonosítottuk a komputationalizmus számos lehetséges téves felfogását, láthatjuk, hogy bizonyos érvek a komputationalizmus ellen szólnak. az elmélet, amely ezekre a tévhitekre épül, nem megalapozott.

Például a következő érv egyértelműen nem engedélyezhető:

P1A komputationalizmus azt mondja, hogy minden mentális művelet számítás.

P2A személyi számítógépem számításokat végez.

A komputationalizmus azt mondja, hogy a személyi számítógépem műveletei mentális műveletek.

P3De a személyi számítógépemnek nyilvánvalóan nincs esze.

A komputationalizmus hamis.

A P1-P3 helyiségek nem vitatottak. P2 és P3 egyértelműen igazak, és a komputationalizmus valóban azt az állítást teszi, amit a P1-ben neki tulajdonítanak. Az érvelés a P1-től és P2-től a köztes következtetésig tartó átmenetben hibás. A következtetés nem

igazságtartó; az időközi következtetés valójában hamis.

Azt, hogy a következtetés nem igazságmegőrző, úgy tudjuk bizonyítani, hogy ellenpéldákat adunk az alkalmazott következtetési formára, mivel az igazságmegőrzés logikai forma kérdése (erről bővebben a 15. fejezetben). A következtetés logikai formájú: Állítsuk A-t úgy, hogy "Melbourne-ben", B-t pedig úgy, hogy "Ausztráliában" (és C és x esetében bármit, amit csak akarunk), és máris világos ellenpéldát kapunk ennek az érvelési formának az érvényességére - egy demonstrációt arra, hogy a premisszák igazsága nem garantálja a következtetés igazságát.

Világosan láthatjuk azonban, hogy a P1 téves értelmezése hogyan vezethet minket a közbenső következtetés levonásához, ha a P2-t vesszük alapul. A komputacionalizmus valóban azt állítja, hogy minden mentális művelet számítás, de a fordítottja, mint láttuk, nem igaz. Következésképpen az a tény, hogy valami számításokat végez, nem garantálja, hogy mentális műveleteket is végez. Ha P1-et tévesen a fordítottjaként olvasnánk (hogy *csak* a mentális műveletek számítástechnikai műveletek - vagyis *minden* komputáció mentális művelet), akkor tévesen azt hihetnénk, hogy a fenti érv egy érvényes formája.

10.1. gyakorlat

Konstruáljon megalapozatlan érveket a komputacionalizmus ellen, amelyek az általunk vizsgált egyéb lehetséges tévhitekre építenek. Minden egyes esetben magyarázza meg, hogy az érv miért nem megalapozott.

Ha van egy hamis premissza (egy állítást a komputacionalistának tulajdonít), magyarázza meg, hogy a premissza miért hamis a komputacionalizmus helyes megértése fényében.

Ha az érvben érvénytelen következtetés szerepel, adjon ellenpéldát az alkalmazott logikai forma érvényességére, és magyarázza meg, hogy egy adott téves felfogás hogyan vezethet ahhoz, hogy valaki azt higgye, hogy az érv érvényes formát képvisel.

10.2 SZOFTVER ÉS WETWARE

A komputacionalizmust gyakran az elme "szoftveres" felfogásaként írják le. Az emberi agyat úgy tekintik, mint ami a biológiai számítási hardvert - vagy *wetware-t* - biztosítja, amely az embereknek az elme képességét adja. E nézet szerint az elme megléte annak függvénye, hogy az embernek a megfelelő program fut-e a *wetware*-ében.

Ez szilárd módszertani keretet biztosít számunkra a mentalitás

vizsgálatához. Mi érdekelhet minket, ha a komputacionalizmus

helyes, a különböző mentális folyamatok programja(i) meghatározása a [LÉLEK] felépítése céljából.

A kognitív diszciplínákban széles körben elterjedt a komputacionalizmus, és nagy szerepet játszik a kutatási programok tájékoztatásában. Az empirikus kognitív tudományágak mindegyike más-más módon közelíti meg a mentalitás vizsgálatának átfogó célját, a saját diszciplináris módszereiknek és feltevéseiknek megfelelően. A komputualizmus iránti elkötelezettség azonban minden esetben széleskörű módszertant biztosít e kérdések kutatásához. A számítási hipotézist valló kutatók célja, hogy olyan beszámolókat készítsenek, amelyek elvileg számítással megvalósíthatók. Vagyis arra törekednek, hogy a mentális folyamatokról mint effektív eljárásokról szóló beszámolókat dolgozzanak ki. Ez nagyon gyakran magában foglalja az informatikusokkal való együttműködést a mentális jelenségek számítási modelljeinek ~~kihívásán~~

Ez a módszertan indította el a *mesterséges intelligencia* néven ismert kutatási hagyományt. A "mesterséges intelligenciának" vannak gyengébb és erősebb értelmezései. A *gyenge mesterséges intelligencia*-kutatási program egyszerűen arra irányul, hogy olyan műtárgyakat építsenek, amelyek képesek bizonyos, az intelligenciát (bár gyenge mértékben) alkotó funkciókat megvalósítani. Ez az a fajta "mesterséges intelligencia", amelyet gyakran használnak fehéráruk eladására.

A *mesterséges intelligencia erős* kutatási programja lényegesen nagyobb érdeklődésre tart számot, és központi szerepet kap ebben a kötetben. A program elkötelezi magát amellel, hogy olyan műalkotások kifejlesztésének lehetőségét kutatja, amelyeknek olyan értelemben *van elméjük*, ahogyan mi magunkat is elmének tekintjük.

Vannak bizonyos mentális képességek, amelyek a jelek szerint csak az emberi elmére jellemzőek. Ezek közé tartozik az a képesség, hogy komplexen és absztrakt módon gondolkodjunk olyan dolgokról, mint a matematika, a logika és a metafizika, valamint a nyelvhasználat. Mind a racionális képesség, mind a nyelvi képesség számos olyan, alacsonyabb rendű kognitív folyamatot is magában foglal, mint például a megkülönböztetés, a tanulás és az emlékezés képessége. Ezeket az alacsonyabb rendű folyamatokat más állatok is többé-kevésbé megvalósítják. A magasabb rendű kognitív funkciók, mint az absztrakt gondolkodás, a nyelvi produkció és a nyelvi megértés azonban egyedülállóan emberi, és mint ilyenek, céljaink szempontjából elsődleges meghatározói annak, hogy milyen intelligenciát tulajdonítunk az embereknek.

Következésképpen a következő fejezetekben a racionális és nyelvi képességekkel rendelkező számítási eszközök fejlesztésének különböző módszereivel fogunk foglalkozni.

Az utolsó szakaszban még többet fogunk mondani arról, hogy milyen feltételek mellett *tulajdoníthatunk* mentalitást egy műtárgynak.

fejezetben, amikor a Turing-tesztet tárgyaljuk. Mielőtt azonban erre rátérnénk, még többet kell mondanunk a komputacionalizmusról. Kezdjük azzal, hogy az eddigi beszélgetésünkből kirajzolhatjuk a komputacionalizmus néhány előnyét.

Először is, és ez a legnyilvánvalóbb, a komputacionalizmus kibővíti a funkcionista keretet. Emlékezzünk vissza a 6. fejezetből, hogy a funkcionista szemléletet kissé hiányosnak találtuk, mint a mentalitás "fekete doboz" szemléletét. A komputacionalizmus elmondja nekünk, hogy mi történik a fekete doboz belsejében, nevezetesen a *számítás*. Következésképpen a komputacionalizmus egyértelmű módszertant biztosít a mentalitás vizsgálatához - arra kell törekednünk, hogy a kognitív képességek számításos elszámolását nyújtsuk.

Másodszor, a komputacionalizmus lehetővé teszi számunkra, hogy az elme és az agy közötti kapcsolatot egy hasznos wetware/software megkülönböztetés alkalmazásával pontosítsuk. E nézet szerint az elme az, amit az agy csinál. Más szóval, az agyak biztosítják a számítási erőforrásokat az [ÉSZ] működtetéséhez.

Ezen a ponton azt is megjegyezhetjük, hogy a komputacionalizmus megőrzi a funkcionista elméletek szubsztrátfüggetlenségét, és következésképpen összeegyeztethető az elme tisztán materiális felfogásával anélkül, hogy az ausztrál materializmust megghiúsító többszörös realizálhatósági kifogásoknak vagy a jelképes fizikalizmus módszertani ürességének áldozatául esne.

Úgy tűnik tehát, hogy a komputacionalizmus élvezi más elméleteknek az erősségeit anélkül, hogy azok legrosszabb gyengeségeinek lenne kitéve. Van azonban számos ellenvetés, amelyet felhozhatunk az ellen a nézet ellen, hogy minden mentális művelet számítás. A következő szakaszokban számos prima facie ellenvetést fogok megvizsgálni - azt a fajtát, amelyet tipikusan a számításelmélet ellen szoktak felhozni első látásra. Minden egyes esetben bemutatom, hogy egy funkcionista hogyan védheti meg az elméletet a kérdéses ellenvetéssel szemben. A bonyolultabb filozófiai érveket a komputacionalizmus ellen a 17. és 18. fejezetre hagyjuk.

10.3 VARIÁCIÓ

Az eltérésből származó kifogás a következőképpen hangzik. A komputacionalizmus azt mondja, hogy az embereknek elméjük van azáltal, hogy [MIND] megvalósítják, de az emberi elmék nagyon változatosak. Hogyan lehetséges ez, ha minden elmét ugyanannak a formális rendszernek az izomorfizmusának tekintünk?

Az biztos, hogy az egyes elmék között jelentős eltérések

vannak. Mindannyian differens hiedelmekkel, vágyakkal, érzelmi reakciókkal és mentális képességekkel rendelkezünk. Egy komputualista természetesen elismeri ezt - tagadni ezt ostobaság lenne. Ez nem válasz a

kifogás, azonban. Ehhez az izomorfizmus fogalmának továbbgondolása szükséges.

A formális rendszerek olyan változókat tartalmaznak, amelyek a rendszer bármelyik példányában értéket kapnak. Ahhoz, hogy két rendszer izomorfikus legyen egymáshoz, *formálisan* ekvivalensnek kell lenniük. Ez azonban nem jelenti azt, hogy egy formális rendszer minden izomorfizmusának minden és csakis ugyanazokat az értékeket kell a változókhoz rendelnie. Valójában ez nagyon ritkán lesz így.

A 8.1. gyakorlat során a formális rendszer [ADD] különböző izomorfizmusait instanciáltad, hogy megállapítsd, milyen függvénynt számol a program. Mindegyik esetben differfikus értékeket rendeltél a kezdeti bemeneti változókhoz.

Továbbá, ha vesszük az [ADD] két izomorfizmusát, amelyek mindegyike azonos kezdeti értéket kapott, majd összehasonlítjuk a regiszterek tartalmát a műveletek differális szakaszaiban, akkor differális értékeket fogunk látni az egyes regiszterekben. Az egyes rendszerek műveletei azonban formailag továbbra is ekvivalensek - továbbra is az [ADD] izomorfizmusai.

Tekintsük most a [MIND] rendszert. Ez egy rendkívül összetett rendszer lehet, nagyon-nagyon sok változóval. Feltehetően a [MIND] minden egyes példánya a működés egy meghatározott szakaszában van. Feltételezhetően a [MIND] minden egyes példánya a változókhoz rendelt értékek (bár talán kissé eltérő) hozzárendelésével kezdődik.

Ezért nem meglepő, hogy a [MIND] két adott példánya nagymértékben különbözik egymástól a rendszer változóihoz jelenleg rendelt értékek tekintetében.

Legalábbis első látásra úgy tűnik, hogy a hiedelmek és a vágyak minden elmében ugyanazt a funkcionális szerepet töltik be. Ez nem jelenti azt, hogy az én hitem, hogy (bármilyen) pontosan ugyanazt a funkcionális szerepet tölti be az én mentális életemben, mint az Ön hitének ugyanez a tartalma az Ön mentális életében. Inkább azt akarom mondani, hogy a hiedelmek, *qua* hiedelmek, funkcionális szerepet játszanak a mérlegelésben, a tervezésben, a cselekvésre való motivációban és így tovább.

Mivel a hiedelmek és a vágyak rendkívül összetett módon lépnek kölcsönhatásba ezekben a mentális funkciókban, nem várhatjuk el, hogy az én hitem, hogy (bármilyen) ugyanazt a cselekvést fogja eredményezni, mint a te hited ugyanarról a tartalomról. Azt azonban el kell várnunk, hogy *kvázi* hiedelemként mindkettőnk hiedelmei működnek a mérlegelésben, tervezésben stb.

Egy komputationalista valószínűleg azzal érvel, hogy a hiedelmek, vágyak és hasonlók *tartalmát* - azt, *amiről* szólnak - a változókhoz rendelt értékek szempontjából kell értelmezni. Az

110

[ADD] rendszerben az R_2 tartalma (bármi legyen is az) mindig ugyanúgy működik - ez az az érték, amely csökken, miközben mi

inkrementum R_1 . Hasonlóképpen, mondja a komputationalista, a [MIND]-ben lévő "hit-doboz" tartalma mindig ugyanúgy fog működni, még akkor is, ha annak tartalma (amiről a hit szól) jelentősen eltérhet a rendszer különböző példányaiban vagy ugyanazon rendszer működésének különböző szakaszaiban.

Következésképpen egy komputationalista tarthatja azt, hogy az elmék a [MIND] izomorfizmusai (funkcionális ekvivalensei), mégis boldogan elismerheti, hogy minden elme különbözik, kisebb vagy nagyobb mértékben, a hiedelmek, vágyak és hasonlók kontenciáját tekintve. A [SZELLEMEK] izomorfizmusai közötti változatosságot is könnyen megmagyarázhatják a sajátos tartalmú hiedelmek, vágyak stb. által részben meghatározott konkrét cselekvésekkel, mivel a cselekvés oksági meghatározóit bonyolultan összefüggőnek fogják tartani sok változó (sok hiedelem, vágy stb. tartalma) értékével. A mentális *tartalom* eszméjéről még sokat fogunk beszélni a 18. fejezetben.

Ez az érvelés nem szól közvetlenül arról a tényről, hogy az elmék a képességek tekintetében is különböznek. Például egyes elmék jobban alkalmazkodnak a formális gondolkodáshoz, egyes elmék jobban tudják használni a nyelvet, egyes elmék képzetebbek a ~~nyelvi~~ ^{mentális} eszközökkel való munkában, és egyes elmék képesek rendkívüli művészeti alkotások létrehozására. Ez elvezet bennünket a következő ellenvetésünkhöz.

10.4 TANULÁS

Az elmék képesek tanulni. Következésképpen a különböző elmék különböző dolgokat tudnak tenni. Más szóval, egyes elmék képesek olyan funkciókat ellátni, amelyeket más elmék nem (vagy bizonyos funkciókat jobban tudnak ellátni, mint a legtöbb más elme). Hogyan tudja ezt a komputationalizmus megmagyarázni?

Láttuk, hogy a komputationalizmus hogyan tud magyarázatot adni az elmék közötti eltérésekre a tartalmuk tekintetében; az elmék azonban a képességeik tekintetében is különböznek. Ez két további kihívás elé állítja a komputationalistákat. Először is, számítással kell magyarázatot adnia az új képességek elsajátítására, azaz meg kell határoznia a tanulást irányító algoritmusokat.

Ideális esetben szeretnénk meghatározni az algoritmus(oka)t (vagy az algoritmusok variációit), amelyek *ténylegesen* irányítják a tanulást az emberi elmében. A komputualizmus e kifogással szembeni védelmében azonban elég, ha megmutatjuk, hogy a tanulás elvileg effektív, és ebben az esetben bármely helyesen működő algoritmus(ok) specifikációja megfelel.

Másodszor, a számítástudósnak meg kell magyaráznia, hogyan lehet funkcionális ekvivalencia két különböző kapacitású rendszer

112
között,

azaz hogyan lehetséges, hogy a [MIND] két izomorfizmusa differens dolgokat *csinálhat*.

Az utóbbi kihívás az alapvetőbb, ezért először erre válaszolok. Gondolkodjunk el azon, hogy mit jelent az, hogy két formális rendszer izo- morfikus egymáshoz. A 7.6. szakaszban azt mondtuk, hogy egy [A] formális rendszer izomorf egy [B] formális rendszerrel, iff ha [B]-t [A]-ból [A]-ból szimbólumok egyenletes helyettesítésével tudjuk levezetni. Ha tehát például veszünk egy sakktáblát valamilyen konfigurációban, és az összes bábut azonos érmeikkel helyettesítjük, akkor az eredmény ugyanannak a formális rendszernek egy másik izomorfizmusa lesz.

Továbbá, ha van két sakktáblánk, és az egyiket több lépésen keresztül játszuk, akkor ugyanannak a formális rendszernek két izomorfizmusa lesz differenciális állapotban. Ekkor az egyik táblán elérhető lehet egy olyan lépés - pl. a sáncolás -, amely a másik táblán nem elérhető.

Az az állítás, hogy két [A] és [B] formális rendszer izomorfikus, a következő: minden x állapotra és R szabályra, ha az R alkalmazása ([A] az x állapotban) y állapotot hoz létre, akkor az R alkalmazása ([B] az x állapotban) y állapotot hoz létre.

Annak megértése, hogy a differens állapotú izomorf formális rendszerek differens kimeneteket eredményezhetnek, ha a szabályalkalmazások azonos sorozatát adjuk meg, lehetővé teszi az elmék közötti eltérések egyszerű magyarázatát egy adott helyzetre adott válaszuk tekintetében. Ha két elme differens hiedelmekkel, vágyakkal stb. rendelkezik, akkor az [ÉTKEZETEIK] differens állapotokban vannak, ezért nem várhatjuk el, hogy ugyanaz a szabályalkalmazás mindkét [ÉTKEZET]-ben ugyanazt a kimenetet eredményezze.

Az elmék közötti eltérések megértéséhez is tettünk némi lépést a képességek tekintetében - differenciális szabályok alkalmazhatók a [MIND] izomorfizmusaira differenciális állapotokban. Itt azonban még többet kell mondani.

Tekintsük az [UM] rendszert a 9. fejezetből. Emlékezzünk vissza, hogy az [UM] úgy működik, hogy dekódolja az R_1 értéket, és lefuttatja az érték által kódolt programot (ha egyáltalán kódol programot). Most nézzünk egy hasonló regiszteres gépet - nevezzük [OS]-nek.

[OS] nagyszámú regisztert különít el, amelyekben saját értékeit tárolja. Nagyszámú regiszter is rendelkezésre áll majd, amelyekben számításokat végezhet ezeken az értékeken. Az [OS] által tárolt értékek egy része algoritmusok (programok) kódja lesz. Ezeket az [OS] más értékeit programbemenetként használva lehet végrehajtani a számításokra fenntartott helyen, és ezek kimenete az [OS] további értékeiként tárolható.

Így például az [OS] egyik regisztere tartalmazhat #[ADD] (az [ADD] program kódja). A működés egy bizonyos pontján a

az [OS] program megcímezheti ezt a regisztert, és két másik regiszter értékét bemenetként használva instanciálja az [ADD]-t, a kimenetet pedig egy másik regiszterben tárolja.

Tegyük fel, hogy az [OS] program számos algoritmus működését irányítja (a regisztereiben tárolt értékek formájában), hogy ezek közül sok lehet működésben egy adott időpontban, hogy az algoritmusok kimenete más algoritmusok bemenetként használható, és hogy az algoritmusok kimenete meghatározhatja, hogy melyik algoritmus lesz a következő (és milyen értékekkel).

Az [OS] program, bár bonyolultabbnak hangzik, mint az [UM], egyértelműen effektív. Amikor arról beszélünk, hogy *egyszerre* sok algoritmus működik, akkor azt értjük, hogy sok algoritmus van folyamatban; azonban minden egyes időlépésnél csak egy algoritmus egy lépése kerül végrehajtásra. Az [OS] tehát egy ugyanolyan regisztergépes program, mint bármelyik másik, és így az [UM] által kiszámítható.

A [MIND] rendszer a [OS] egy nagyon összetett változataként értelmezhető. Úgy működik, hogy számos olyan algoritmus működését szabályozza, amelyek egyenként mentális funkciókat látnak el - algo- ritmusok, amelyek az érzékszervi adatokat érzékelési reprezentációkká alakítják, algoritmusok, amelyek a testi mozgásokat szabályozzák, algoritmusok, amelyek a nyelvi produkciót és megértést szabályozzák, algoritmusok, amelyek a hiedelmeken, vágyakon és hasonlókon alapuló cselekvéseket határozzák meg és így tovább.

A szoftveranalógiával élve, a [MIND] leginkább egyfajta *operációs rendszerként* értelmezhető, amely nagyszámú alkalmazás egymással szorosan összefüggő műveleteit kezeli, és ellenőrzi a hardvereket, amelyekben a szoftverek instancializálódnak.

A [MIND] által alkalmazott algoritmusok egy része *tanulási algoritmusként* szolgál. *Tanuló algoritmusnak* nevezünk egy algoritmust, ha alkalmazása eredményeként a rendszer nagyobb képességekkel ruházódik fel.

Legalább kétféleképpen lehet egy elmét nagyobb képességekkel felruházni: új információkhoz juthat, vagy pedig készségeket vagy képességeket szerezhet, illetve fejleszthet. Az előbbi esetben az elme megtanulja, *hogy a dolgok így vannak*. Az utóbbiban az elme megtanulja, *hogyan* történnek a dolgok.

A [MIND] rendszerben a dolgok megismerése megfelel az új tartalom tárolásának - azaz új értékek tárolásának a regiszterekben (ahol ezek az értékek kódolják a tartalmat). A dolgok hogyanjának megtanulása megfelel új algoritmusok tárolásának vagy a meglévő algoritmusok optimalizálásának.

Tekintettel arra, hogy a [MIND]-et egyfajta [OS]-ként értelmezzük, amely folyamatos iterációs működésben van, és amely bővíthető - azaz

olyan algoritmusokkal rendelkezik, amelyek tartalmat tárolnak, és amelyek további algoritmusokat generálnak vagy optimalizálnak, amelyek mindkettő nagyobb kapacitást biztosít a rendszernek -, a következőkben

segíthetünk magunknak egy, a komputalista számára elérhető magyarázatot adni az [ÉN]-ek közötti eltérésekre a kapacitások tekintetében.

Tekintettel arra, hogy - amint azt már megállapítottuk - bármely két [MIND] nagy valószínűséggel különböző állapotban van, és tekintettel arra, hogy a működésének időtartamától és a kapott bemenetek jellegétől függően egy [MIND] több vagy kevesebb tárolt tartalmat és több vagy kevesebb (jobban vagy rosszabbul optimalizált) tárolt algoritmust tartalmaz, nem meglepő, hogy bármely két [MIND] jelentősen különbözik egymástól a kapacitások tekintetében.

Ez a magyarázat választ ad a tanulási ellenvetés által a komputeralisták számára felvetett két kihívás egyikére - megmagyarázza, hogy két izomorf formális rendszer hogyan rendelkezhet differens képességekkel, és így hogyan lehet az elméket a [MIND] izomorfizmusainak tekinteni annak ellenére, hogy az elmék között óriási eltérések vannak a tekintetben, hogy mire képesek.

A számítástudomány számára azonban továbbra is kihívást jelent a tanulást irányító algoritmusok meghatározása. Erre a kihívásra a következő fejezetek különböző pontjain visszatérünk, különösen a 13. fejezetben, amikor az automatizált gondolkodási rendszereket tárgyaljuk, és a 19. fejezetben, amikor a mesterséges neurális hálózatokban történő tanulást vizsgáljuk. A 16. fejezetben azt is megvizsgáljuk, hogy az emberek hogyan tanulnak nyelveket, és megvizsgáljuk a bizonyítékokat arra, hogy ez a tanulás szabályokkal szabályozott.

Most két ellenvetést vizsgáltunk meg, amelyeket a komputationalizmus ellen felhozhatunk. Ezek lényegében ugyanannak az ellenvetésnek az erősebb és gyengébb változatai - az elmék változnak. Mindkét esetben láttuk, hogyan válaszolhat egy komputalista ésszerűen. Nézzünk meg egy további ellenvetést azzal az elmélettel szemben, hogy az elmék komputációs eszközök.

10.5 KREATIVITÁS

Egy másik szokásos prima facie kifogás az emberi alkotóképességre hivatkozik, a következőképpen. A formális rendszerek működése teljesen mechanikus, de az elmék *kreatívak*. Az elmék nagyszerű művészeti, zenei, építészeti és irodalmi alkotásokat hoznak létre, és hatalmas innovációs képességgel rendelkeznek. Az emberi elmére jellemző kreativitás meggyőző bizonyítéknak tűnik a komputationalizmus ellen, amely a mentalitást tisztán mechanikus műveletekkel próbálja magyarázni.

Minden bizonnyal úgy tűnik, hogy semmi sem áll távolabb egy algoritmikus folyamattól, mint egy műalkotás megfestése vagy egy zenekari szimfónia komponálása. Amint azonban a 2. fejezetben láttuk, a dolgok látszólagos volta nem megbízható mutatója annak, hogy a dolgok milyenek.

A komputeralista számára a kihívás itt az, hogy megmagyarázza, hogyan lehet az általunk paradigmátikusán "kreatívnek" nevezett mentális funkciókat algo-ritmikusan, az intuícióval ellentétesen megvalósítani. Ehhez a magyarázathoz szükségünk van a kreativitás fogalmának megértésére.

A komputualizmus ellenzői a kreativitás egy olyan definícióját támogathatják, amely a következőképpen hangzik: *egy tevékenység akkor kreatív, ha az eredménye egy olyan alkotás (műalkotás, kompozíció stb.) létrehozása, amelyet nem lehetett volna egyszerűen egy szabályozott eljárás követésével létrehozni.*

Bár ez a definíció némileg intuitív - végül is mindannyian eléggé biztosak vagyunk benne, hogy da Vinci nem számok alapján festett, amikor a *Mona Lisát* festette -, a kérdést meglehetősen egyenesen a számítástudósok ellenében teszi fel. Pontosan arról van szó, hogy a kreativitás algoritmikusan számon kérhető-e vagy sem.

Mi az tehát, ami a kreativitásban olyan, hogy a kezdeti intuíciónk az, hogy a kreatív viselkedést szembeállítjuk a szabályokkal szabályozott viselkedéssel?

Nos, először is, nem mindenki egyformán kreatív. Az emberek különböző képességekkel rendelkeznek a kreatív vállalkozásokban való részvételre. Azt azonban már láttuk, hogy a komputualista hogyan tudja figyelembe venni a képességek tekintetében fennálló eltéréseket, így ez nem elégséges ellenvetés, de a helyes irányba mutat.

Úgy tűnik, hogy ami kezdetben ellenzi, hogy a kreatív viselkedés szabályozott legyen, az az intuíció, hogy ha szabályozott lenne, akkor könnyebben tanítható lenne. Ez jellemző azokra, akiket alkotó mesterekként dicsérünk - művészek, kézművesek, zeneszerzők stb. - hogy tehetségükben van valami rejtélyes mások számára. Továbbá úgy tűnik, hogy amikor kreatív törekvésekről van szó, az ember vagy "rendelkezik vele", vagy nem. Bizonyos, hogy az ember megtanulhat különböző technikákat és módszereket az anyagokkal való munkához, hogy bizonyos effektusokat hozzon létre; az azonban nem világos, hogyan tanulhatna meg valaki "kreatívnek lenni" önmagában.

Ez azonban nem zárja ki annak lehetőségét, hogy az ilyen viselkedés mögött valóban számítási folyamatok állnak. Természetesen a komputualizmus ellenfelei számon kérhetik, hogyan lehet egy ilyen viselkedést számítással megvalósítani; azonban bizonyos válaszok rendelkezésre állnak.

A komputationalista például elmondhat egy történetet arról, hogy bizonyos algoritmusok megvalósításukhoz bizonyos számítási erőforrásokra van szükség, így az egyes algoritmusok elsajátításának képessége tekintetében mutatkozó eltérések az [ÉN]-ek megvalósításának alrendszerében (agyában) lévő

eltérésekkel magyarázhatók.

Mindenesetre a komputationalistáknak joguk van ahhoz, hogy ne válaszoljanak arra a kérdésre, *hogy miként érhető el a kreatív viselkedés komputacionálisan?* - amíg a "kreatív viselkedés" fogalma nem válik

szigorúbban meghatározott. A kreativitás olyan, egymástól független, plauzibilis magyarázatának hiányában, amely közvetlenül ellenzi az ilyen viselkedés szabályvezérelt voltának lehetőségét, a komputeralista nem áll rosszabbul ebben a tekintetben, mint más teoretikusok.

A számmisztikával szembeni számos előzetes ellenvetést megvizsgáltunk, és láttuk, hogy az ellenvetés minden esetben megbukik. Még meg kell vizsgálnunk az elmélet elleni kifinomultabb érveket. Ezekkel addig nem foglalkozunk, amíg nem láttunk néhány olyan ~~rossz~~ intelligencia-alkalmazást, amely a filozófiai ellenvetések kontextusát adja.

Összefoglalva, megvitattuk a komputacionalizmus néhány gyakori félreértését és néhány rossz érvet az elmélet ellen, amelyek ezekre a félreértésekre építenek. A komputacionalizmus egy sajátos változatát terjesztettük elő, amely szerint az elme egy számítógépes operációs rendszerrel analóg, amely irányítja az alkalmazások működését és ellenőrzi a hardvert, amelyben ez a rendszer működik (a központi idegrendszert). Tanúi voltunk annak is, hogy a komputacionalizmussal szembeni számos, első látásra felmerülő ellenvetés kudarcot vallott.

Ebben a fejezetben az utolsó dolog, mielőtt továbblépnénk a különböző mesterséges intelligencia-alkalmazások vizsgálatára, hogy eltöltünk egy kis időt azzal, hogy elgondoljuk azon, hogy milyen feltételek mellett vagyunk hajlandók mentalitást tulajdonítani.

10.6 ATTRIBÚCIÓS MENTALITÁS

Hogyan határozzuk meg, hogy van-e valaminek szellemi élete vagy sem? Amikor nem vagyunk a filozófiai szkepticizmus szorításában, a legtöbben meg vagyunk győződve arról, hogy a körülöttünk élők mentális élettal rendelkeznek - legalábbis szokás szerint úgy viselkedünk, mintha így lenne. Tekintettel arra, hogy csak a saját mentális állapotainkhoz van privilegizált hozzáférésünk, mi az, ami más emberekkel kapcsolatban arra készlet bennünket, hogy elhiggyük, hogy van elméjük?

Nos, először is, a többi ember fizikailag nagyon hasonlít ránk, és ugyanolyan lénynek tűnik, *mint* mi. Ezért elvárjuk tőlük, hogy különböző tulajdonságaik legyenek velünk, például hogy hasonló mozgásra és a környezettel való interakcióra legyenek képesek. Ez kiterjed arra az elvárásra is, hogy a többi embernek is van belső mentális élete. Nyilvánvaló azonban, hogy az ilyen fizikai hasonlóság nem *elégséges az általunk* tapasztalt mentalitáshoz, mivel ezeket a feltételeket a nem emberi állatok és a nagyon

kisgyermek is teljesíthetik, akiknek megítélésünk szerint nincs olyan gazdag mentális életük, mint amelyet mi élvezünk.

Valójában a fizikai képességek hasonlósága sem *szükséges* a mentalitás tulajdonításához, mivel elképzelhetünk valakit, akinek egyértelműen van elméje, abban az erős értelemben, ahogyan mi magunkat tartjuk elmének, de mégsem rendelkezik ezekkel a fizikai képességekkel.

Milyen feltételeknek felelnek meg tehát más emberek, hogy mentalitást tulajdonítsunk nekik? A kulcs itt - ami nem meglepő, tekintve a következő fejezetek állítólagos fókuszát - a nyelvhasználatunkban rejlik.

A hozzáértő és kifinomult nyelvhasználat a mentális jellemzője. A "kompetens" és "kifinomult" alatt itt azt értem, hogy a beszélő képes újszerű kijelentéseket felfogni és produkálni, különböző nyelvi eszközöket használni különböző effektusok elérése érdekében, és különböző témákról beszélni, amelyek az absztrakció különböző fokozatait foglalják magukban.

Az ilyenfajta nyelvi képesség bizonyítása révén - ami, ha szabad megjegyezmem, racionális képességet is magában foglal - vagyunk hajlandóak mentalitást tulajdonítani azoknak a dolgoknak, amelyek így *viselkednek*. Még akkor is, ha ez a képesség nagyon nem szabványos módon valósul meg, úgy tűnik, készek vagyunk ezt a mentalitás tulajdonításához szükségesnek tekinteni. Senki sem fogja például tagadni, hogy Stephen Hawkingnak van elméje, méghozzá kiváló elméje.

Itt óvatosnak kell lennünk - én a szükségességre vonatkozó állítást tettem, de nem a szükségességre vonatkozó állítást. Az afáziának (nyelvi deficitnek) vannak olyan esetei, hogy az érintettek a nyelvi képességek bármely standard tesztjén elbuknak, de más tekintetben mégis eléggé alkalmasak, olyannyira, hogy ~~lax~~ szilárd mentalitást tulajdonítanánk nekik. Erre a pontra akkor térünk vissza, amikor a 16. fejezetben az emberi nyelvet tárgyaljuk.

A mentalitás nyelvi képességének ez a szufficienciája az elme jelenlétének durva és kész tesztjét adja számunkra. Ha a kísérleti alany képes megfelelni bizonyos általunk meghatározott feltételeknek, amelyek célja, hogy megvizsgáljuk az alany nyelvhasználati képességét, akkor azt gondolhatjuk, hogy ez elegendő ahhoz, hogy - legalábbis munkahipotézisként - feltételezzük, hogy az alannak valóban van elméje.

A való életben ez az értékelés egy folyamatos vállalkozás. Hajlamosak vagyunk alapértelmezésben feltételezni, hogy másoknak van elméjük, és a mentális képességeikről alkotott becslésünket a bizonyított szellemi kapacitásuknak megfelelően felülvizsgáljuk. Társadalmi interakcióinkon való némi elmélkedés is ezt bizonyítja, így nem fogok tovább érvelni mellette. Természetesen az is igaz, hogy a nyelv segítségével vagyunk képesek mások elméjét vizsgálni (itt a "vizsgálni" szó laikus

értelemben értendő). A legegyszerűbb módja annak, hogy megtudjuk mások hitét, vágyait és így tovább, ha beszélünk velük (bár ez természetesen hibás).

Pontosan ezt az intuíciót ragadta meg Alan Turing a *Computing Machinery and Intelligence* (*Számítógépek és intelligencia*) című korszakalkotó cikkében, melyben

Turing elhatározta, hogy megvizsgálja, milyen feltételek mellett vagyunk hajlandóak mentalitást tulajdonítani egy műtárgynak.

Turing felállított egy tesztet, amelyet *Turing-teszt* néven ismerünk, és amelynek sikeres teljesítését javasolta, hogy a tesztalany elmével rendelkezzen.

A *Turing-tesztet* gyakran gyengébb formában találják, mint ahogyan azt Turing elképzelte, ezért alkalmazzunk némi óvatosságot a felállításánál. A teszt egy régi társasjáték, az *imitációs játék* kiterjesztése. Az imitációs játékokban egy férfit és egy nőt külön szobában helyeznek el, egy harmadikban pedig egy kérdezőbiztos tartózkodik. A kihallgató egy közvetítő naplót használ, és jegyzetek átadásával (amelyeket a közvetítő felolvas a kihallgatónak) meghatározott ideig kihallgatja a férfit és a nőt. A kihallgatónak lehetősége van arra, hogy a férfit és a nőt bármilyen általa megfelelőnek ítélt témában kikérdezze. A játék lényege, hogy a kihallgató megpróbálja megállapítani, melyik szobában van a férfi és melyikben a nő, és hogy a férfi és a nő megpróbálja becsapni a kihallgatót.

Turing azt javasolta, hogy ennek a játéknak az adaptációja az artefaktumok (számítógépek) mentalitásának barométereiként szolgálhatna. A számítástechnikai eszközökkel szemben nem szabad elvárnunk, hogy megfeleljenek a fizikai hasonlóság feltételeinek, vagy képesek legyenek bizonyos fizikai feladatok elvégzésére ahhoz, hogy azt mondhassuk, hogy van elménk. Hiszen láttuk, hogy ez sem nem szükséges, sem nem szufficiens. Azt azonban elvárhatnánk, hogy ha egy tárgy képes kielégíteni bennünket nyelvi kompetenciájával, akkor ez nagyon jó jele annak, hogy a tárgynak van elméje.

A Turing-teszt az imitációs játékot - valószínűleg már nyilvánvaló módon - úgy alakítja át, hogy egy műtárgy nyelvi képességének értékelésére tisztességes eszközt nyújtson. A következőképpen zajlik. Az egyik szobában elhelyezünk egy számítógépet, a másikban egy embert, a harmadikban pedig egy emberi kérdezőt. A kérdező képes kommunikálni a számítógéppel és az emberrel egy billentyűzetten és monitoron keresztül. A kérdezőnek meghatározott idő áll rendelkezésére, hogy minden résztvevőt bármilyen általa megfelelőnek ítélt témában kikérdezzen, és ezalatt megpróbálja megállapítani, hogy melyik beszélgetőpartnere az ember, miközben az ember és a számítógép is megpróbálja meggyőzni őt arról, hogy ő az ember. Ha a kijelölt idő végén a kérdező nem tudja megkülönböztetni a gépet az embertől, akkor azt kell mondanunk a gépről, hogy van elméje.

Mik a kezdeti reakciói a Turing-tesztre, mint a gépi intelligencia mutatójára? Milyen problémákat vet fel

elképzелhető, hogy ezt a módszert használjuk az elme jelenlétének megállapítására?

Ne feledjük, hogy az állítás *nem* az, hogy a Turing-teszt teljesítése *elegendő* az elme *meglétéhez*. A gondolat az, hogy a Turing-teszt teljesítése *jó okot* ad arra, hogy feltételezzük, hogy a tesztalany mentális étellel rendelkezik. Jelen formájában a Turing-teszt alul specifikált marad. Még nem határoztuk meg a teszt elvégzésének időtartamát, és sok múlik ezen. Ha például a tesztet csak öt percig kell elvégezni, akkor joggal lehetnek kétségeink a megbízhatóságát illetően, mint az emberi életképesség mutatója. Ha viszont a kihallgatónak megengedik, hogy annyi időt töltsön el, amennyit csak akar, amíg úgy érzi, hogy képes megfelelő értékelést készíteni, akkor úgy tűnik, hogy sokkal nagyobb bizalmunk lehet az értékelésben - különösen, ha, mint ahogyan azt tennünk kell, az értékelést úgy vesszük, hogy a további ilyen vizsgálatok eredményeitől függően felülvizsgálható.

1991 óta évente megrendezik a Loebner-díj (a filantróp szponzorról elnevezett) versenyt, amelyen a résztvevők a Turing-teszt formalizált változatának vetik alá magukat. A verseny egy napon át tart, és 100 000 dolláros díjjal, valamint aranyéremmel jutalmazza azt a gépet, amely a bírák szerint megkülönböztethetetlen az embertől. Eddig még egyetlen gép sem közelítette meg ezt. Minden évben azonban 2000 dolláros díjat és bronzérmeket kap az a gép, amelyik a legjobban teljesít a nevezők közül. Turing 1950-ben azt tervezte, hogy 2000-re a gépek képesek lesznek meggyőzően teljesíteni egy közepes időtartamú Turing-tesztet. Kétségtelen, hogy csalódott lenne a jelenlegi helyzet miatt - a Loebner-díjra pályázókkal folytatott beszélgetések leirataiból kiderül, hogy néhány meglehetősen nem túl finom kérdésfeltevés a gép megkülönböztetésére szolgál. résztvevők nagyon rövid időn belül.

Ez azonban nem *ipso facto* a Turing-tesztet, hanem a mesterséges intelligencia jelenlegi helyzetét vádolja.

Lehet, hogy egy ilyen viselkedési teszt, bár bizonyos erős feltételek mellett jó indikátor, egyszerűen nem oldja meg kielégítően a kérdést a problémás esetekben. Ilyen esetekben nem nyilvánvaló, hogy milyen más módszereket alkalmazhatnánk az elme jelenlétének vagy hiányának megállapítására.

10.3. gyakorlat

Vajon a Turing-teszt egy olyan változata, amely több bíró részvételével és korlátlan időtartamban zajlik, a mentalitás megbízható mutatója lenne? Tudsz más feltételeket is

128

mondani.

vagy korlátozások, amelyeket beépíthetünk a tesztbe, hogy növeljük annak megbízhatóságát?

Most álljunk meg, és gondolkodjunk el azon, hogy hol vagyunk és hová tartunk. Ebben a könyvben eddig bemutatunk különböző filozófiai problémákat, amelyek az elme megértésével kapcsolatosak, és megvizsgáltunk néhány elméletet, amelyek válaszokat javasolnak ezekre a kérdésekre. Bemutattunk néhány kezdetleges neuroanatómiai ismeretet is, hogy megértsük, hogyan működik az agy, mivel nyilvánvalóan fontos kapcsolat van az agy és az elme között.

Az általunk tárgyalt filozófiai elmeelméletek közül a funkcionalizmus volt a legkevésbé problematikus és a leginkább kielégítő. Azonban azt hagyta bennünk, hogy többet szeretnénk tudni a szóban forgó funkciókról. Annak érdekében, hogy jól felkészüljünk a válaszadás egy sajátos módjának megértésére, ezután szigorúan kidolgoztuk a számítás pontos beszámolóját.

Ezzel a felismeréssel felvértezve, most a funkcionalizmus kibontásának egy sajátos módját vizsgáltuk meg, amely feltételezi, hogy a szóban forgó funkciók számítások - nevezetesen a komputacionalizmust. Megfontoltuk azt a kérdést is, hogy milyen megfelelő feltételek mellett kellene mentalitást tulajdonítanunk. Erről még sokat fogunk beszélni, ahogy haladunk előre.

Az elkövetkező fejezetekben látni fogjuk, hogy a komputualizmus iránti elkötelezettség hogyan ad módszertant a mesterséges intelligencia kutatásához. Különböző számítási problémákra fogunk összpontosítani, amelyeket meg kell oldani ahhoz, hogy egy eszközt felruházzunk a gondolkodás és a nyelvhasználat képességével.

Útközben olyan bizonyítékokat fogunk gyűjteni, amelyek a komputationalizmus tarthatóságát érintik - erre a kérdésre kifejezetten visszatérünk a 17. fejezet elején, amikor elkezdünk komoly kihívásokat és kifinomult filozófiai ellenvetéseket felhozni az elmélettel szemben.

Itt az ideje, hogy elkezdjük elmesélni a mesterséges intelligencia történetét, amely a *keresés* fogalmával kezdődik.

11. FEJEZET

KERESÉS

Az általunk ismertett komputationalista hipotézis alapján a mentalitás vizsgálata a formális rendszer [MIND] és alkotóelemei működésének vizsgálatát jelenti.

A mesterséges intelligencia klasszikus vagy *szimbólumrendszeres* megközelítése a [MIND] kognitív funkcióinak algoritmusait próbálja meghatározni, és a hozzájuk kapcsolódó formális rendszereket vizsgálja.

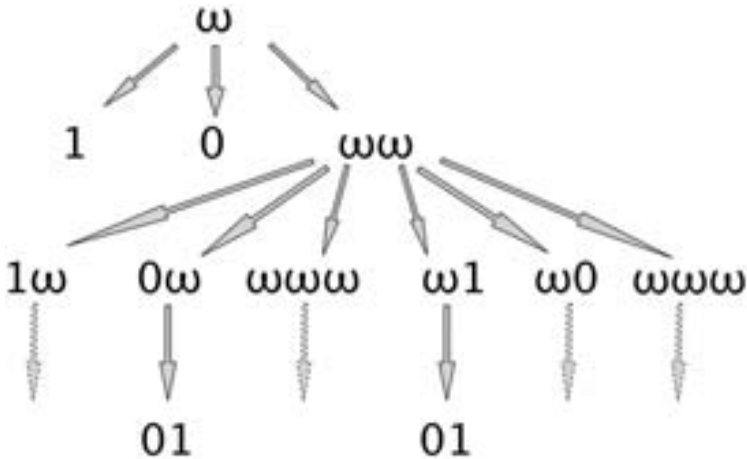
A klasszikus mesterséges intelligencia hagyományában számos probléma annak meghatározására redukálódik, hogy egy bizonyos állapot generálható-e egy adott formális rendszerben vagy sem, vagy egy rendszer egy adott generált állapotának megtalálására. Más szóval, számos mesterséges intelligencia probléma egy adott állapot keresésének problémájára redukálódik.

Ebben a fejezetben röviden végigvesszük a formális rendszerek generációs fájának keresésére szolgáló különböző módszereket. A 7.5. fejezetben felfrissíthetjük a generációs fákkal és a kapcsolódó terminológiával kapcsolatos ismereteinket.

11.1 FELÜLRŐL LEFELÉ, ALULRÓL FELFELÉ

Egy formális rendszer egy állapotának meghatározására, illetve a rendszer egy adott állapotára vonatkozó levezetés megtalálására az egyik módszer a rendszer teljes generációs fájának megkonstruálása. Ha a minket érdeklő állapot megjelenik a fán, akkor az egy generált állapot, és a levezetés(ek)et a gyökércsomópontig visszamenőleg követve olvashatjuk el.

Tegyük fel például, hogy a 7.5. szakaszban szereplő [BIN] rendszert vizsgáljuk, és az érdekel bennünket, hogy a 01 állapotot generálja-e. A [BIN] rendszer generációs fáját a kezdeti állapottól kezdve felépíthetjük, és végigmegyünk az összes lehetséges módon, ahogyan a szabályokat az egyes csomópontokra alkalmazni lehet. A 11.1. ábra a [BIN] generációs fáját mutatja, amelyből kiderül, hogy (és hogyan) generálódik a 01 állapot.



11.1 ábra Top-down keresés.

Az ilyen jellegű keresést *felülről lefelé irányuló* keresésnek nevezzük. Gyakran azonban effitikusabb, ha a levezetni kívánt állapotból indulunk ki, és visszafelé haladva a szabályokon keresztül megnézzük, hogy vissza tudunk-e jutni a kiindulási állapothoz. Ezt a fajta keresést *alulról felfelé irányuló* keresésnek nevezzük. A 11.2. ábra egy alulról felfelé irányuló keresést ábrázol a [BIN]-ben a 01-es állapotra.

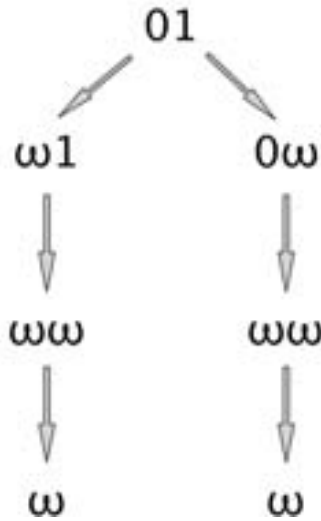
A felülről lefelé irányuló keresés egyik előnye, hogy ~~az~~ Minden generált állapot megjelenik egy befejezett top-down fán. Az alulról felfelé irányuló keresésnek viszont az az előnye, hogy az összes és csakis a lehetséges levezetéseket szolgáltatja a megoldási állapothoz, ha vannak ilyenek.

Annak eldöntésében, hogy a felülről lefelé vagy alulról felfelé irányuló keresés lesz-e hatékonyabb, a kapcsolódó *elágazási tényező* az egyik fő szempont. A fa elágazási tényezője - az egyes csomópontok alatti leszármazottak átlagos száma - határozza meg a fa bonyolultságát, így ideális esetben minimalizálni szeretnénk. Amint a 11.1. és 11.2. ábra mutatja, egy rendszerben a felülről lefelé irányuló keresés elágazási tényezője gyakran különbözik az alulról felfelé irányuló keresés elágazási tényezőjétől. A [BIN] esetében az alulról felfelé irányuló keresés effitikusabbnak bizonyul.

11.1. gyakorlat

Mi határozza meg, hogy egy alulról felfelé haladó fa elágazási tényezője alacsonyabb vagy magasabb lesz-e egy rendszer esetében, mint egy felülről lefelé haladó fa

elágazási tényezője?



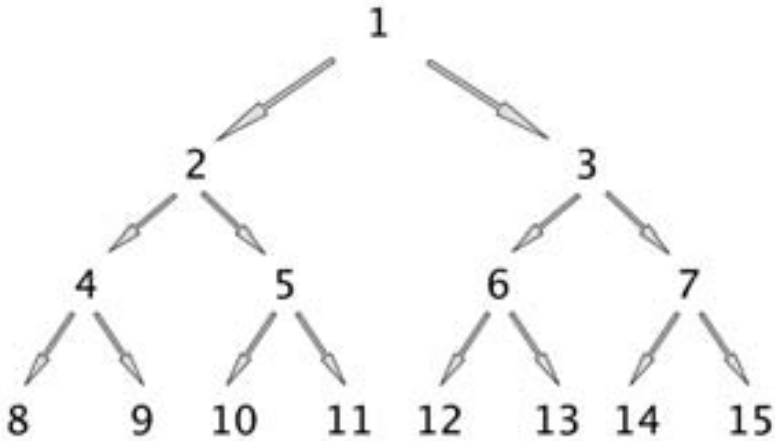
11.2. ábra Alulról felfelé irányuló keresés.

11.2 SZÉLESSÉG KONTRA MÉLYSÉG

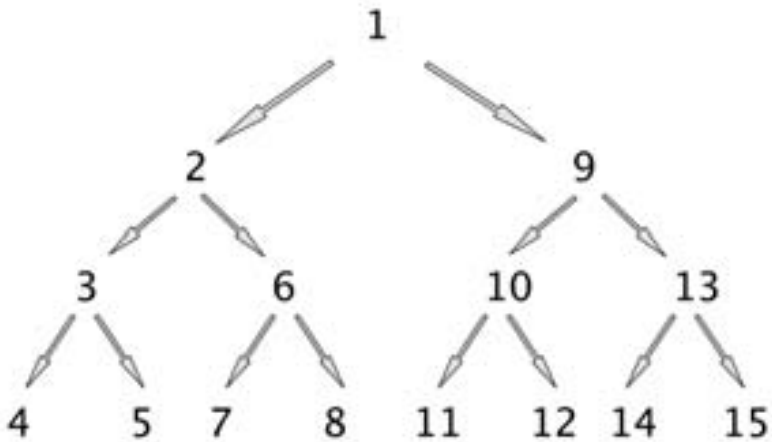
Miután döntöttünk a felülről lefelé és az alulról felfelé történő keresés között, döntenünk kell a fa keresésének eljárásáról. Tartsuk szem előtt, hogy bármely formális rendszer, amely a mesterséges intelligencia kutatói számára érdekes lehet, lényegesen összetettebb lesz, mint az itt vizsgált játékpéldák. Következésképpen a megfelelő keresési eljárás kiválasztása jelentős hatással lehet a keresés elvégzéséhez szükséges számítási erőforrásokra. Sőt, mint látni fogjuk, ez jelentheti a különbséget a sikeres keresés és a soha le nem álló keresés között.

Az *első szintű* keresés során egy adott szinten lévő összes csomópontot kimerítően átkutatjuk, mielőtt a következő szintre ereszkednénk. A 11.3. ábra azt mutatja, hogy milyen sorrendben keressünk a fa csomópontjait, ha a keresést balról balra haladva, szélesség szerint végeznénk.

Az első keresés egyértelmű előnye, hogy kimerítő. Ha a fán bárhol találunk megoldási állapotokat, a keresés megtalálja azokat. A szélesség első keresés hátránya azonban, hogy a szükségesnél sokkal több számítási költséggel járhat. Ez különösen akkor van így, ha a megoldás állapota messze van az egyik ágban.



11.3 ábra Széleskörű keresés.



11.4. ábra Mélységi első keresés.

Ha arra számítunk, hogy számos megoldási állapot lesz az ágak hosszú szakaszán, akkor célszerűbb a *mélységi* keresés. A mélységi keresés során az ágak teljes hosszában végig kell keresni, amíg vagy találunk egy megoldási állapotot, vagy egy végállapotot. Ha elérünk egy terminális csomópontot, akkor csak addig megyünk vissza, amíg le nem tudunk ereszkedni egy keresetlen ágon.

A 11.4. ábra azt mutatja, hogy milyen sorrendben keresnék a 11.3. ábrán látható fát, ha mélységben először, balra először keresnék. Míg a mélység szerinti keresés jelentősen hatékonyabb lehet olyan helyzetekben, amikor számos megoldási állapot van a fa ágainak mélyén.

egy fát, a használata veszélyt rejt magában. Ha a fa bármelyik ága végtelen lesz - ahogy az érdekes formális rendszerekben gyakran előfordul -, és mélység szerinti keresést használunk, akkor fennáll annak a veszélye, hogy a keresésünk soha nem ér véget. Ha egy végtelen ágon nincs megoldási állapot, akkor a mélység szerinti keresésünk a végtelenségig folytatódik az ágon.

Hasznos stratégia lehet a mélység- és a szélesség alapú keresés *kombinálása*. Ha például tudjuk, hogy a fa egyes ágai, amelyek minket érdekelnek, a végtelenbe nyúlnak, és gyanítjuk, hogy a fában számos megoldási állapot lesz a csomópontok között elosztva, jóval lejjebb, akkor a mélység szerinti keresést *csak egy bizonyos szintig* végezhetjük. Ha az adott szintig kimerítjük a csomópontokat anélkül, hogy találnánk egy megoldási állapotot (ami egyenlő azzal, hogy a szélesség első keresést végeztük el az adott szintig), akkor folytatjuk a mélység első keresést egy mélyebb szinten.

11.3 HEURISZTIKUS KERESÉS

A "Breadth first" és a "Depth first" mindkettőt *vak* keresésnek nevezzük - a keresett csomópontok *megoldáshoz való közelségének* figyelembevétele nélkül végzik.

Gyakran előfordul azonban, hogy rendelkezünk valamilyen hatékony eljárással, amellyel meghatározhatjuk, hogy egy adott csomópont leszármazottai között milyen valószínűséggel vannak megoldási állapotok. Az a függvény, amely egy ilyen algoritmust alkalmaz a csomópontokra, és ennek megfelelően értéket rendel hozzájuk, *heurisztikus függvény*.

A jó heurisztikák meghatározása nagyon gyakran a legnehezebb eleme a szimbolikus mesterséges intelligencia hagyományában szereplő komplex problémák megoldásának.

Vegyünk például egy klasszikus mesterségesintelligencia-projektet: egy algoritmus meghatározása a jó sakkozashoz. A sakkozás egyik heurisztikus megoldása az lenne, ha egyszerűen megalkotnánk a sakk generációs fáját, majd bármely adott játékállapothoz - amelyet a fa egy csomópontja képvisel - a győztes állapotokat reprezentáló leszármazó csomópontok számának megfelelően értéket rendelnénk. A játékban ezután csak annyit kell tennünk, hogy mindig arra a pozícióra lépünk, amelynek a csomópontja a legjobb értékkel rendelkezik a rendelkezésre álló értékek közül.

Egy olyan szabály esetén, amely felső határt szab a megengedett lépések számának (például az ötven lépés szabálya), a fát végesnek tarthatjuk, így ez az eljárás megvalósíthatónak tűnik, amíg nem kezdjük számszerűsíteni a sakk pusztá komplexitását.

A sakk generációs fája drasztikus kombinatorikus robbanástól

szerved. Csak a második szinten négyszáz csomópont van. Mire a tizedik iterációhoz érünk - amikorra minden játékosnak már

csak öt lépést tett - körülbelül kilencmilliárd csomópont van. A huszadik iterációnál valami 1030 csomópont nagyságrendű, és ez *nagyon* nagy szám.

Még ha másodpercenként százbillió állapotot is tudnánk generálni, akkor is körülbelül tízmilliárd évbe telne, hogy csak a sakk első tíz fordulójának generációs fáját létrehozzuk.

Nyilvánvaló tehát, hogy az általunk leírt naiv heurisztika számítási szempontból tarthatatlan. Hasonlóképpen, a vak keresési technikák - bár megvannak a maguk hasznai - gyakran számításilag tarthatatlanok a mesterséges intelligencia kutatói számára érdekes, megfelelő bonyolultságú formális rendszerek esetében. Az exponenciálisan összetett formális rendszerek generációs fájának vizsgálata csak olyan okos heurisztikák segítségével lehetséges, amelyek jelentősen csökkentik a számítási terhelést.

A sakkjátékot megfelelően jól irányító heurisztikus függvények meghatározása a legjobb emberi játékosok kihívásához rendkívül nehéznek bizonyul, és e függvények kiszámítása jelentős számítási erőforrásokat igényel. Ezt a következő fejezetben tárgyaljuk tovább, amikor a játék automatizált módszereit szembeállítjuk az emberi módszerekkel.

Egy heurisztikus függvény birtokában még mindig el kell döntenünk, hogy a heurisztikus értékek hogyan irányítsák a keresést. Az egyik ilyen heurisztikus keresési eljárás a *hegymászás*.

A hegymászó keresés során minden egyes csomópontnál kiértékeljük a közvetlen leszármazott csomópontok heurisztikus értékét, majd a legjobb heurisztikus értékkel rendelkező csomópontra lépünk. Ha elérünk egy végcsomópontot anélkül, hogy találnánk egy megoldási állapotot, akkor visszamegyünk a következő legjobb lehetőséghez.

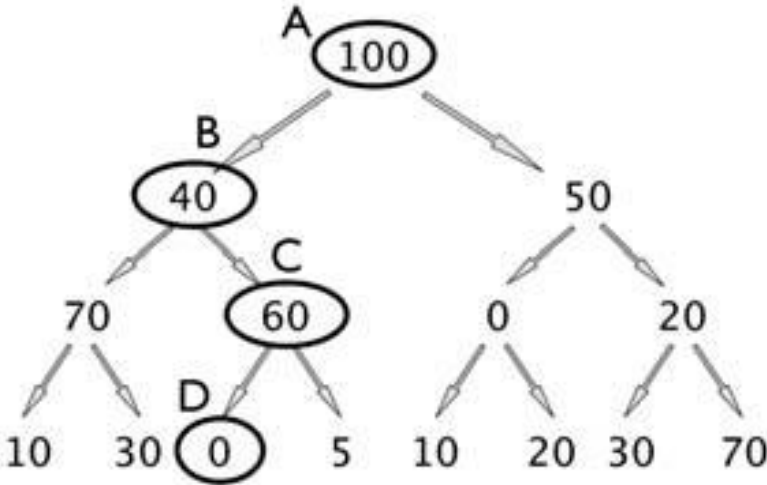
A 11.5. ábrán látható betűk azt a sorrendet jelölik, amelyben az ábrázolt generációs fát keresnénk, ha a heurisztikánk az ábrán látható módon értékeket rendelne a csomópontokhoz, ahol a 0 a megoldás állapotát jelenti.

11.2. gyakorlat

Ha a 11.5. ábrán látható fára egyszerűen csak a szélesség első, balra első keresést alkalmaznánk (heurisztikus függvény alkalmazása nélkül), hány csomóponton kellene áthaladnunk, mielőtt elérnénk ugyanazt a megoldási állapotot? Mi lenne, ha mélységben először, balra először keresést alkalmaznánk? Mi lenne, ha mélység első, jobb első használnánk?

Amint a 11.2. gyakorlat mutatja, ez a heurisztikus keresés - a

keresett csomópontok számát tekintve - jelentősen hatékonyabb, mint bármelyik vak keresési eljárásunk. Az is igaz azonban, hogy *csomópontonként* számításigényesebb, mivel a meghatározás mellett



11.5. ábra Hegymászó keresés.

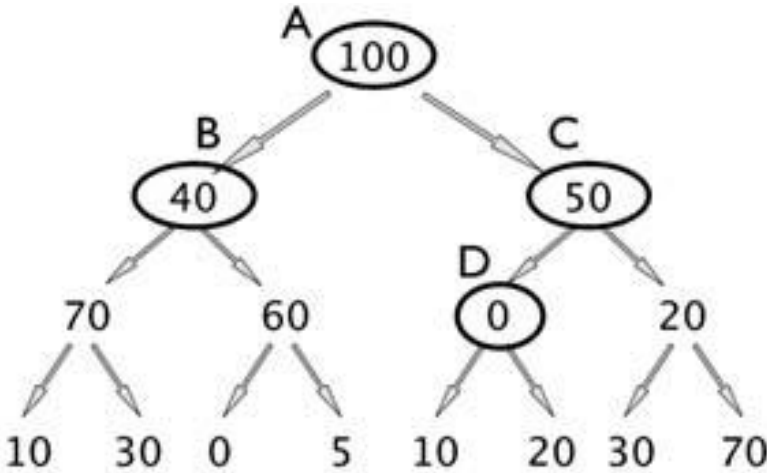
hogy minden egyes keresett csomópont egy megoldási állapotot képvisel-e, szintén egy heurisztikus függvényt alkalmazunk.

Bármely, akár mérsékelt komplexitású generációs fa esetében azonban a heurisztikus függvény alkalmazásának csomópontonkénti számítási költségét messze felülmúlja a heurisztikus útmutató által biztosított általános számítási megtakarítás, mivel jellemzően jelentősen csökkentjük a vizsgálandó csomópontok számát.

Sajnos azonban a heurisztikák csak azok - heurisztikák. Nem döntési eljárások. Más szóval, bár a heurisztikus értékek valamilyen módon jelzik egy adott csomópont megoldáshoz való közelségét, a hozzárendelt érték nem feltétlenül tükrözi jól a csomópont *tényleges* megoldáshoz való közelségét. A jobb heurisztikák pontosabb jelzéseket adnak, de mindig előfordulhat, hogy a heurisztikus értékek tévútra vezetnek minket.

A hegyászó keresés tévútra vezethet bennünket, ha egy adott keresési útvonal kezdetben távolodik a megoldás állapotától (a heurisztikánk által mért kritériumok szerint), de aztán visszalendül, és gyorsabban vezet a megoldáshoz, mint más keresési utak.

Hogy ezt világossá tegyük, gondoljunk a következő térbeli analógiára. Tegyük fel, hogy az esőerdőben vagyok nagyjából 100 méter magasan, és le akarok jutni a tengerpartra, amelyről tudom, hogy nagyjából keletre van tőlem. Két ösvényt látok, amelyek közül az egyik északkelet felé, a másik pedig délkelet felé vezet. A fák között nem sokat látok egyik ösvényből sem, de azt látom, hogy az északkeleti ösvény finoman vezet.



11.6 ábra A legjobb első keresés.

lefelé, míg a délkeleti útvonal kezdetben egy kis emelkedő megmászásával jár. A céloim tengerszinten van, és a lehető legkevesebb effortot akarom ráfordítani, ezért az északkeleti utat választom. Kiderül azonban, hogy az északkeleti út kanyargós, és útközben sok emelkedőt kell megmásznom. A délkeleti út viszont a kezdeti emelkedő után gyorsan leereszkedik a partra, és végig csökken a szintkülönbség.

A leírt forgatókönyvben a hegymászó keresés miatt a megoldáshoz vezető leggyorsabb keresési útvonalról lemaradok, mivel ez az útvonal kezdetben hátrányosnak tűnt. Az ilyen helyzetek kezelésének egyik módja a *legjobb első* keresés.

A legjobb első keresés olyan, mint egy hegymászó keresés, kivéve, hogy a csomópontok kiválasztása nem korlátozódik kizárólag a közvetlen leszármazott csomópontokra. Az eddig keresett csomópontok heurisztikus értékeit is nyilvántartjuk, és a közvetlen leszármazott csomópontok és e többi csomópont közül választunk.

Ez azt jelenti, hogy ha a keresés során olyan csomóponthoz érünk, ahol a heurisztika szerint az összes közvetlen leszármazott csomópont távolabb van a megoldástól, mint egy korábban keresett csomópont, akkor visszamehetünk a fán felfelé, és a következő, jobban értékelt csomópontból kereshetünk.

A hegymászáshoz hasonlóan a legjobb első keresés a legkisebb ellenállás útját választja. A differencia csak annyi, hogy bármelyik csomópontból kibővítjük a rendelkezésre álló választási lehetőségeket - már nem követeljük meg, hogy a keresési utak mindig egy ágon vezessenek le a következő iterációhoz.

A 11.6. ábra azt mutatja, hogy a 11.5. ábrán látható fát a legjobb

124

első keresés alapján hogyan keressünk.

Amint láthatjuk, a legjobb első keresés használata lehetővé teszi, hogy ebben az esetben egy rövidebb levezetést fedezzünk fel a megoldási állapotunkhoz - egy korábbi iterációnál találunk egy megoldási csomópontot.

A legjobb első keresés egyértelműen az általunk vizsgált leghatékonyabb keresési eljárás. Világosnak kell azonban lennie, hogy csomópontként ez a legszámításigényesebb is. Nemcsak egy heurisztikus függvényt kell alkalmaznunk a csomópontokra, hanem szükségünk van a memóriaterületre is, amelyben az összes eddig keresett csomópont értékét rögzíthetjük. A fa összetettségének növekedésével ez a számítási költség jelentős lehet. Még mindig igaz azonban, hogy a keresendő csomópontok számának számítási megtakarítása messze meghaladja a heurisztikus eljárás számítási költségét.

A heurisztikákról még sok mindent el tudnánk mondani, ezért a következő fejezetben azt vizsgáljuk meg, hogyan alkalmazhatnánk a heurisztikus keresést egy klasszikus mesterséges intelligencia alkalmazásban - a játékokban.

12. FEJEZET

JÁTÉKOK

Most, hogy a keresési eljárások és a heurisztikus funkciók alapvető megértésével rendelkezünk, itt az ideje, hogy ezt a megértést a játék automatizált módszereinek megfontolása során alkalmazzuk.

Ez belépési pontot biztosít számunkra annak szélesebb körű vizsgálatához, hogy miként alkalmazhatunk formális rendszereket az intelligenciát alkotó funkciók megvalósítására.

Összehasonlítjuk ezeket a formális játékmódszereket azzal a reflexív megértésünkkel, ahogyan az emberek ilyen játékokat játszanak, mielőtt a következő fejezetekben részletesebb és megalapozottabb összehasonlításra térnénk át az érvelésben és a nyelvben szerepet játszó kognitív funkciók tekintetében.

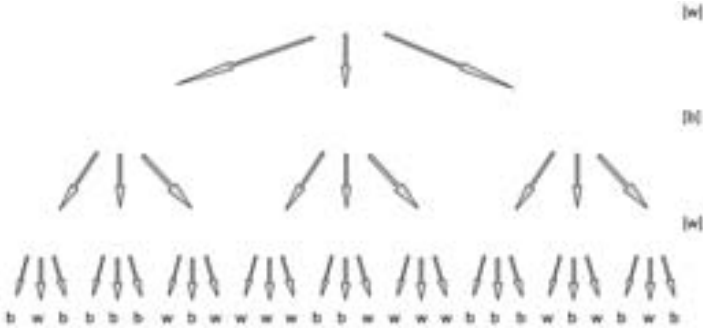
12.1 EGY EGYSZERŰ JÁTÉK

Ha egy játék nagyon egyszerű, nagyon könnyen automatizálhatunk egy ~~játék~~ játékhoz. Elég triviális például egy olyan algoritmus meghatározása a tic-tac-toe játékhoz, amelynek követése mindig vagy győzelmet vagy döntetlent eredményez.

Ilyen esetekben csak annyit kell tennünk, hogy felépítjük a játék generációs fáját, majd a végcsomópontoktól visszafelé haladva meghatározzuk a stratégiát.

Tegyük fel például, hogy van egy játék, amelyet a következőképpen játszunk. Először is kiválasztunk egy kiindulási pozíciót. Ebből a pozícióból és minden további pozícióból pontosan háromféleképpen léphet a játékos. A játékosok feldobnak egy érmét, hogy ki kezd, majd az első játékos lép, a második játékos lép, és az első játékos tesz egy utolsó lépést. E három lépés után a játék olyan pozícióban ér véget, amely az egyik játékos egyértelmű győzelmét jelenti.

A játék generációs fája könnyen felépíthető. Három iterációval és három elágazási faktoriall rendelkezik, így a következő lesz



12.1. ábra Egy egyszerű játék.

huszonhét terminális csomópont, amelyek mindegyike az egyik (fehér) vagy a másik (fekete) játékos győzelmét jelenti.

A 12.1. ábrán látható generációs fa feltételezi, hogy választottunk egy kiindulási pozíciót a játékhoz, és minden végponthoz hozzárendeli azt a játékos, akinek az adott pozíció nyerő pozíció. Nem kell ismernünk a játék vagy a játékmenet részleteit - itt csak a formája érdekel minket. Csupán szemléltetésképpen határoztam meg, hogy melyik csomópont melyik játékos számára jelent győzelmet.

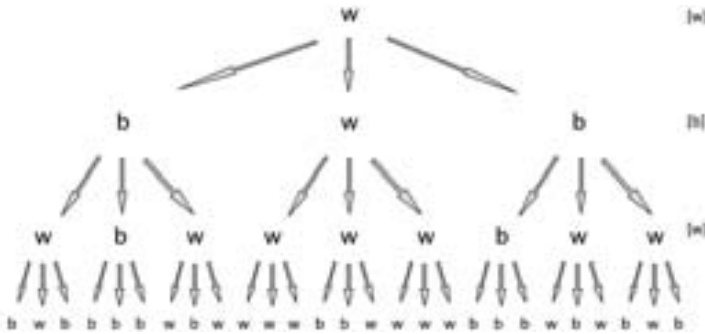
A 12.1. ábrán látható információk birtokában visszamehetünk a fán felfelé, és meghatározhatjuk, hogy a kiválasztott kiindulási pozíció kinek a számára lesz nyerő.

Tegyük fel, hogy például fehér nyeri a dobást, és ő teszi meg az első lépést. Ez azt jelenti, hogy a terminális csomópontok előtti utolsó iteráción fehér lép. Az adott iteráción egy csomópont akkor számít fehér győzelemnek - mivel ez a fehér lépése -, ha van legalább egy olyan leszármazott csomópont, amely fehér győzelmet eredményez, ellenkező esetben fekete győzelemnek számít.

Más szóval, ha fehér nyerő állásba tud lépni, és ő van soron, akkor az állás már nyerő állás fehér számára. Ha azonban az egyetlen lehetséges lépés olyan állásokba történik, amelyek fekete számára nyerő állások, akkor az állás már fekete számára nyerő állás.

Hasonlóképpen, az előző iterációban a fekete kerül sorra. Az adott iterációban egy csomópont akkor számít a fekete győzelmének, ha van legalább egy olyan leszármazott csomópont, amely a fekete győzelmének számít, egyébként pedig a fehér győzelmének számít.

A 12.2. ábra ezt az eljárást alkalmazza az ábrán látható fára. 12.1 és megmutatja, hogy ha fehér nyeri a dobást, akkor fehérnek nyerő stratégiája van. Ez azonban nem jelenti azt, hogy a fehér megnyeri a



12.2. ábra A fehér győztes stratégiája.

játék - elvégre egy emberi játékos is hibázhat. Csupán azt kell mondanunk, hogy van egy algoritmikus *nyerő stratégia*, amely a fehérek számára elérhető.

12.1. gyakorlat

Tegyük fel, hogy a fekete nyeri a dobást, és a 12.1. ábrán látható generációs fát az általunk leírt eljárás szerint töltsük ki. Van-e ebből a kezdeti kombinációból elérhető nyerő stratégia a fekete számára, ha a fekete lép először?

A 12.1. feladat megoldása azt mutatja, hogy a leírt játékban, a 12.1. ábrán látható végpontok hozzárendelését eredményező kiindulási helyzet mellett, a dobást megnyerő játékosnak egy nyerő stratégia áll rendelkezésére. Ez nem biztos, hogy a játék más kezdeti pozíciói esetén is így van, de a nyertes játékosok terminális csomópontokhoz való példakijelölésének szándékos jellemzője.

12.2. gyakorlat (Kihívás)

Tekintsük a tic-tac-toe generációs fáját. Tegyük fel, hogy a nyerő pozíciókat képviselő összes csomópont terminális.

- Hány csomópont van az első iterációnál?
- Melyik iterációnál generálódnak először a terminális csomópontok?
- Hány csomópont van összesen a fán?
- Hány végcsomópontot tartalmaz a fa?
- Hány ilyen csomópont képvisel győztes állapotot bármelyik játékos számára?

Bár ez az eljárás, amely a végpontoktól visszafelé haladva dolgozza fel a nyerő stratégiákat, csak a következő esetekben alkalmazható

csak a legegyszerűbb játékok. Amint azt az előző fejezetben láttuk, az érdekesebb és összetettebb játékok **gyűis** fái exponenciálisan nőnek, ami egy ilyen eljárást számítási szempontból tarthatatlanná tesz.

Ahhoz, hogy meghatározzuk a nyerő stratégiák keresésének módszereit az olyan játékokban, mint a sakk, heurisztikákat kell alkalmaznunk.

12.2 MINIMAX

A 11.3. szakaszban láttuk, hogy egy csomópontra olyan heurisztikus függvényt alkalmazhatunk, amely a rendszer szempontjából releváns kritériumok alapján értékeli az adott csomópont megoldáshoz való közelségét. Egy kétfős játékban azok az állapotok számítanak megoldási állapotnak, amelyek az egyik játékos (tetszőleges) győzelmét jelentik. Ebben az esetben a játék a legmesszebbre kerülhet a megoldási állapottól, ha a másik játékos győzelmét jelenti.

Következésképpen a heurisztikus értékeink azt mutatják, hogy a játék állapota hogyan áll a játékosok lehetséges győzelmének szempontjából. A példa kedvéért tegyük fel, hogy az alacsonyabb számok a fekete, a magasabb számok pedig a fehér győzelméhez való közelséget jelentik.

Tehát a 11.3. szakaszban látott példával ellentétben - ahol a gyökércsomópont a lehető legtávolabb volt a megoldástól - a két játékosra vonatkozó játék gyökércsomópontja pontosan a tartomány közepén fog heurisztikus értéket felvenni (feltételezve, hogy a kezdeti pozíció nem előítéletes egyik játékos javára sem).

A heurisztikus függvények a csomópontokat a csomópontban reprezentált állapot bizonyos belső jellemzői alapján értékelik. Így például a sakkban figyelembe vehetjük az anyagi előnyt, a központ dominanciáját és bizonyos bábuk előretörését, hogy olyan értéket hozzunk létre, amely az adott állapot jóságát tükrözi az egyes játékosok számára. Azonban még egy nagyon jó heurisztikus függvény is korlátozott abban, hogy csak az állapot *belső* jellemzőit veszi figyelembe.

Ezt állítsuk szembe az előző szakaszban vizsgált eljárással. A további iterációk során generált állapotokról szóló információk felhasználása az előzetes csomópontok értékeléséhez olyan eljárás, amely az állapotokon *kívüli* jellemzőket vesz figyelembe.

Nagymértékben növelhetjük a stratégiák meghatározására szolgáló módszer pontosságát komplex játékokban, ha a belső és külső jellemzők kombinációját használjuk útmutatóként. Más szóval, kombinálhatjuk egy heurisztikus függvény használatát egy olyan módszerrel, amely a kiértékelt csomópontokból visszafelé

haladva határozza meg a korábbi csomópontok értékét.

Láttuk, hogy a sakk teljes generációs fáját nem tudjuk megvalósítani. Amit azonban megtehetünk, az az, hogy csak néhány lépést keresünk előre, és egy heurisztikus függvényt alkalmazunk a keresési horizonton lévő csomópontokra. Ezután egy *minimax* eljárást használhatunk arra, hogy a keresési horizonttól visszafelé haladva meghatározzuk a kiértékelt csomópont értékét.

Tekintettel arra a korábbi kikötésünkre, hogy az alacsonyabb heurisztikus értékek a fekete győzelemhez való közelségét jelzik, a fekete egy *minimalizáló* és a fehér egy *maximalizáló* lesz. Más szóval, a fekete mindig olyan állapotokba igyekszik elmozdulni, amelyek alacsonyabb heurisztikus értékekkel rendelkeznek, a fehér pedig olyan állapotokba igyekszik elmozdulni, amelyek magasabb értékekkel rendelkeznek.

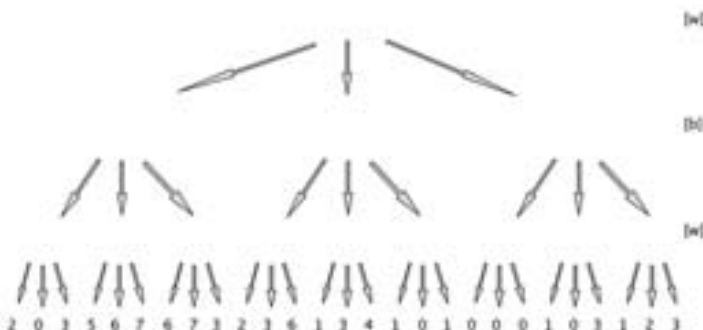
Végezzünk egy kisebb módosítást az előző részben leírt egyszerű játékon, hogy a játék már ne mindig az egyik játékos győzelmével záruljon a három forduló végén. Most tegyük fel, hogy kiválasztunk egy kiinduló pozíciót, és *csak* az első három lépésnél rajzoljuk fel a generációs fát, majd egy heurisztikus függvényt alkalmazva megkapjuk a 12.3. ábrán a *horizontcsomópontoknál* ábrázolt értékeket.

A horizontcsomópontok heurisztikus értékeinek felhasználásával egy *minimax* eljárást alkalmazhatunk a kiválasztott kezdeti pozíció értékének meghatározására, az alábbiak szerint.

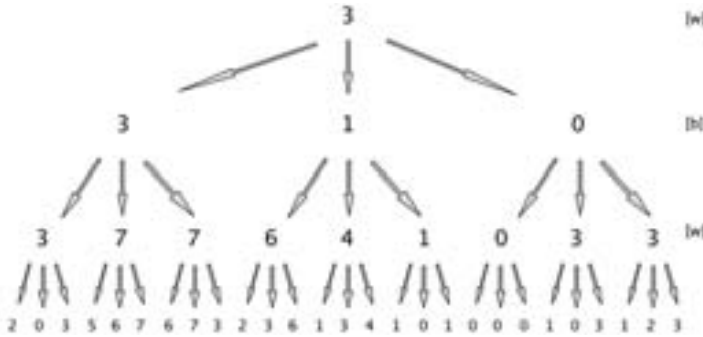
Tegyük fel, hogy a fehér nyeri a dobást, amely esetben a keresési horizontot közvetlenül megelőző iterációban a fehéren a sor, hogy lépjen. Fehér egy *maximalizáló*, így egy csomópont értéke az adott iterációban a közvetlen leszármazott csomópontok értékeinek *maximuma* lesz.

Az ezt közvetlenül megelőző iterációban fekete következik. A fekete egy *minimalizáló*, így az adott iterációban egy csomópont értéke a közvetlen leszármazott csomópontok értékeinek *minimuma* lesz.

Ezzel az eljárással a horizontcsomópontoktól visszafelé haladva határozzuk meg a gyökércsomópont értékét, amint azt a 12.4. ábra mutatja.



12.3 ábra Minimax.



12.4. ábra Győztes stratégia a fehérre.

12.3. gyakorlat

Tegyük fel, hogy a fekete nyeri a dobást, és - a 12.3. ábrán látható horizontcsomópontok azonos értékeit használva - alkalmazzunk minimax eljárást a kezdő pozíció értékének meghatározására.

12.3 VÁGÁS

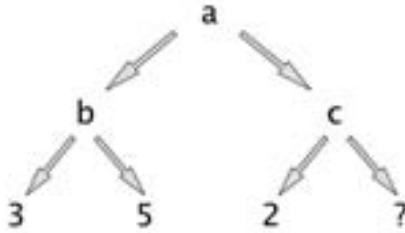
Most már rendelkezünk egy olyan eljárás alapjaival, amellyel egy olyan összetett játék stratégiáit keressük, mint a sakk. Olyan messze előre keresünk, amennyire számításilag lehetséges, a keresési horizont csomópontjaira heurisztikus függvényt alkalmazunk, majd egy minimax eljárással visszamegyünk a keresett csomóponthoz.

Ez az eljárás azonban még mindig a kombinatorikus ~~függ.~~ függ. Szerencsére vannak további lehetőségek a számítási erőforrásaink hatékonyságának maximalizálására a keresés mennyiségének csökkentésével, illetve a keresési fa *metszésével*.

Az egyik ilyen módszer az, hogy a keresés során már felfedezettek alapján megállapítjuk, hogy bizonyos ágakat figyelmen kívül hagyhatunk. A 12.5. ábrán látható fadarab egy olyan (erősen leegyszerűsített) helyzetet mutat, amelyben ez lehetséges.

Mivel a fekete egy minimalizáló, tudjuk, hogy a b csomópont értéke 3 lesz. Azt is tudjuk, hogy a c csomópont értéke *legfeljebb* 2. Következésképpen már tudjuk, hogy az a csomópont értéke 3 lesz (mivel a fehér egy maximalizáló), függetlenül a kérdőjellel jelölt csomópont értékétől.

A példa esetében ez természetesen triviális metszés. Világosnak kell azonban lennie, hogy az elv lényegesen többre is kiterjed



12.5 ábra A fa metszése.

összetett esetekben, és lehetővé teheti számunkra, hogy a keresési fa nagy részét figyelmen kívül hagyjuk, annak alapján, hogy mit fedeztünk fel korábban.

A metszés másik módszere az, hogy olyan vezérelveket találunk, amelyek azt mondják, hogy bizonyos lépéseket egyszerűen ne vegyünk figyelembe. Az összetett játszmák esetében gyakran előfordul, hogy egy adott állapotból elérhetőek bizonyos legális lépések, amelyeket senki sem tenne meg, vagy egyáltalán nem valószínű, hogy bárki is megtenne. Ha tájékozódunk arról, hogy ezek a lépések milyen valószínűtlenek, akkor a keresési fa c csomópontokból kiinduló ágait elvethetjük.

Egy további metszési módszer bizonyos lépéssorozatok memóriából történő lejátszását foglalja magában, anélkül, hogy bármilyen keresést végeznénk. Valójában a számítógépek jellemzően pontosan így játszanak sakkmegnyitásokat. Hozzáférésük van a szabványos megnyitások és variációk nagy adatbázisához, és a kezdeti lépéseket egy előre meghatározott sorrendből játsszák ki.

Minél jobban meg tudjuk metszeni a keresési fát, annál effektívebben használjuk fel a számítási erőforrásainkat. Következésképpen ahhoz, hogy egy gépet úgy programozzunk, hogy jól játsszon egy összetett játékot, a legjobb heurisztikákat kell alkalmaznunk, amelyeket ki tudunk találni, a keresési feladat egyszerűsítésére szolgáló különböző módszerekkel együtt.

12.4 EMBER KONTRA SZÁMÍTÓGÉP

Most egy olyan tevékenységgel foglalkoztunk, amelyet általában az intelligencia alkotóelemének tekintenek - azzal a képességgel, hogy jól tudunk (legalábbis megtanulunk) játszani egy olyan összetett játékot, mint a sakk. Láttuk - legalábbis nagy vonalakban -, hogyan lehet szimbólumrendszereket és keresési módszereket alkalmazni az ilyen játékok stratégiai játékának automatizálására.

A sakk esetében ez magában foglalja a nyitási stratégiák nagy adatbázisát, egy optimalizált heurisztikus függvényt a játékállapotok kiértékelésére, különböző elveket a valószínűtlen

lépések figyelmen kívül hagyására, valamint jelentős számítási erőforrások felhasználását a másodpercenkénti nagyon nagy számú állapot kiértékelésére. Miközben ezt írom, a gépek, amelyek

amelyeket a sakknagymesterek kihívására használnak, másodpercenként körülbelül kétszázmillió állapotot képesek kiértékelni.

Hasonlítsuk össze az automatizált sakkjátékról tanultakat azzal, amit intuitíve tudunk az emberi sakkjátékról.

Először is, az emberek - legalábbis azok, akik nagyon jól játszanak - nagyszámú nyitási stratégiát és azok variációit is megjegyzik. Másodsor, az emberek nagyon jók abban is, hogy meghatározzák, mely lépéseket nem valószínű, hogy egy jól játszó ellenfél megteszi. Azok az emberek, akik nagyon jártasak a sakkban, nagyon jól érzékelik azt is, hogy egy adott játékállapot mennyire jó számukra.

Úgy tűnik tehát, hogy jól illeszkedik az, amit az automatizált sakkjátékról tudunk, és az, amit a gyakorlott emberek játékaról meg tudunk érteni. Vannak azonban fontos különbségek a kettő között, amelyek közül a legnyilvánvalóbb a sakkozó számítógépek által végzett keresés és kiértékelés mennyisége.

Nyilvánvaló, hogy a sakknagymesterek nem alkalmaznak *explicit módon* egy heurisztikus függvényt másodpercenként mintegy kétszázmillió állapotra, hogy értékeljék az aktuális játékállapot jóságát és meghatározzák a stratégiát. Itt azonban nagyon óvatossá kell lennünk, mert ennek fényében kísértésbe eshetünk, hogy egy nagyon gyenge érvet fogadjunk el a komputualizmus ellen.

A nagyon gyenge érvelés a következőképpen hangzik. A számítógépeknek másodpercenként kétszázmillió állapotot kell átkutatniuk ahhoz, hogy a legjobb emberi játékosokkal felvegyék a versenyt. Ezeknek az emberi játékosoknak azonban nyilvánvalóan nem. Tehát az emberi játékosok nem formális rendszereket működtetnek a sakkolás során. Ezért a komputationalizmus hamis, mivel van legalább egy olyan ~~kognitív~~ feladat, amelynek funkciója nem számolható el a formális rendszerek műveleteivel.

Számos oka van annak, hogy ez az érv miért rossz. A legbeszédesebb kritika az, hogy különbséget teszünk az emberi teljesítmény és a legjobb próbálkozásaink között, amelyekkel megpróbáljuk megteremteni azt.

Könnyen lehet, hogy az emberi játékosok valóban alkalmazzák ezeket a módszereket - és ez így is lesz, ha a komputeralizmus igaz -. Csakhogy az embereknek nagyon jó vezérlő elveik vannak, amelyek lehetővé teszik számukra, hogy lényegesen kevesebb keresést folytassanak a nyerő stratégiák meghatározásához. Az a tény, hogy ezek az emberi játékosok nem képesek *kifejezni*, hogy milyen elvek vezérlik őket, nem szól az ellen, hogy kognitív funkcióik a formális rendszerek működésével magyarázhatók. Elvégre mentális életünk nagy része átláthatatlan az önvizsgálat számára. Ez csupán empirikus vizsgálataink legjobb eredményei

ellen szól, amikor e rendszerek működésének meghatározására törekszünk.

A komputacionalizmus elleni érvelés másik kritikája, hogy egyenlőségelet tesz az általunk kifejezetten és tudatosan végzett kognitív funkciók és a kognitív funkciók között. Ha azt mondjuk, hogy egy nagymester nem mérlegel *explicit módon* kétszázmillió állapotot másodpercenként, az nem egyenértékű azzal, hogy az értékelési kognitív folyamatai *implicit módon* nem egyenlők ilyen számú állapot mérlegelésével. Az utóbbira vonatkozó állításra támaszkodni, amikor csak az előbbi bizonyítható, olyan kétértelműség, amely a komputeralistával szemben felveti a kérdést.

Bár nyilvánvalóan óvatosságnak kell lennünk a számítógépes és az emberi sakkjáték közötti különbségtételből levont következtetésekkel kapcsolatban, vannak további említésre méltó különbségek.

Egyrészt a minimax eljárás feltételezi, hogy az ellenfél mindig a számára optimális lépést teszi meg. Nagyon gyakran előfordul azonban, hogy az emberi játékos nem optimális lépést tesz. Előfordulhat, hogy nem ismeri fel az optimális lépést, vagy előfordulhat, hogy szándékosan játszik szuboptimális lépést, hogy megpróbálja rákényszeríteni ellenfelét, hogy egy bizonyos stratégiát kövessen.

A számítógép természetesen az is előfordulhat, hogy nem optimális lépést tesz, egyszerűen azért, mert nem értékeli helyesen. Az ellenfél stratégiájának effektálása érdekében szándékosan nem optimális lépést tenni azonban teljesen más dolog.

Ahhoz, hogy valaki szokatlan lépéssel aktívan próbálja befolyásolni az ellenfele stratégiáját, meg kell értenie, hogyan gondolkodik az ellenfele a játékról. Más szavakkal, szükség van arra, hogy az ellenfél bizonyos körülmények között valószínűleg hogyan fog lépni. Ezt úgy lehet a legjobban megismerni, ha minél többet játszunk a szóban forgó ellenfél ellen.

A számítógép és az ember közötti versenyek során a programozók általában módosíthatják a számítógépes programot a játékok között, hogy az emberi játékos által alkalmazott stratégiákat ellensúlyozzák. Fontos megjegyezni, hogy ahol a számítógépnek emberi beavatkozásra van szüksége ahhoz, hogy képes legyen reagálni egy adott játékos stratégiájának sajátosságaira, ott a gyors reagálású alkalmazkodás a sajátosságokra olyan dolog, amiben az ember természeténél fogva kiválóan teljesít. Ez elvezet bennünket az emberi és a számítógépes játék közötti utolsó ponthoz, amelyet itt megvizsgálunk.

Gondoljunk arra a tényre, hogy az emberi sakkállapotok értékelését vezérlő elvek sokkal jobbak, mint a számítógép által alkalmazott elvek, mivel az összehasonlíthatóan jó játékhoz lényegesen kevesebb sakkállapot kifejezett értékelésére van szükség.

Tekintsük a gyors

egy adott játékos stratégiájának sajátosságaira való érzékenység, amelyet az emberi játékosok mutatnak, de a számítógépek nem.

Úgy tűnik, hogy van valami, ami mindkét képességet szolgálja, amiben az emberek nagyon jók, de amit egyáltalán nem említettek a gépi játékról szóló vitában. Az emberek kiválóan képesek a környezeti ingerekből *mintákat kivonni*, és felismerni, ha ezek a minták ismét előjönnek, még akkor is, ha csak részben jönnek elő, vagy ha a mintában van némi eltérés.

Ez egy fontos felismerés, amelyet a 19. fejezetben még hosszasan fogunk tárgyalni. Egyelőre jegyezzük meg, hogy ez nem *ipso facto* érv a komputualizmus ellen, de magyarázati terhet ró a komputualistára - a teher az, hogy számítással magyarázza ezt a kognitív képességet.

Most, hogy legalább kezdetlegesen összehasonlítottuk azt, ahogyan az emberek komplex játékokat játszanak, azzal, ahogyan a számítógépek is képesek ilyen játékokat játszani, itt az ideje, hogy figyelmünket az emberi racionális és nyelvi képességekre fordítsuk.

13. FEJEZET

GÉPI GONDOLKODÁS

Ebben a fejezetben a racionális képesség - az *érvelési* képesség - vizsgálatát kezdjük el. Az automatizált reasoning módszerek alapos áttekintése külön köteteket igényelne. Az automatizált érvelés egy olyan fajtájára fogunk koncentrálni, amely jól megfelel céljainknak - a *szakértői rendszerek* tervezésére.

Megnézzük, hogyan lehet formális rendszerek segítségével újraalkotni egy adott terület emberi szakértőjének érvelési folyamatait.

Mielőtt ezt megtennénk, először is tisztáznunk kell néhány fogalmat és terminológiát, amelyek a logika tanulmányozásával kapcsolatosak.

13.1 LOGIKA ÉS KÖVETKEZTETÉS

Az első dolog, amit meg kell értenünk, hogy különbség van a logika és a logika között. A logika olyan kutatási hagyomány, amelynek vizsgálati tárgyai a logikák. Ezek a logikák formális rendszerek, és nagyon sok van belőlük.

A logikák célja a *következetességi* vagy *logikai következményi* viszonyok formális kódolása. Más szóval, a logika olyan formális rendszer, amely módszereket biztosít annak meghatározására, hogy mi miből *következik*, mint *logikai forma* kérdése.

Egy másik fontos pont, hogy mint minden formális rendszer, a logika is csak formális tulajdonságokkal foglalkozik. Bár a logikák elemei *értelmezhetők* úgy, hogy jelentenek valamit, a jelentéssel kapcsolatos kérdések soha nem befolyásolják a logikai következmények meghatározását. Más szóval, az, hogy valami logikailag következik-e valamilyen más dologból, vagy sem, *teljes mértékben* a vonatkozó logikai formák kérdése. Ennek jelentősége a fejezet későbbi feladataiban fog kiderülni.

Nem várom el, hogy még nem érti pontosan, mi is az a logikai forma, vagy hogyan lehet megkülönböztetni a logikai formákat. Nem áll szándékomban

a logikába való teljes körű bevezetés, ezért a további tárgyalást ezzel kapcsolatban a következő szakaszra tartogatom, ahol néhány egyszerű példát fogunk látni.

Mivel a logikák formális rendszerek, el kell gondolkodnia azon, hogy mik az állapotok és mik a szabályok. A mi céljainkra az állapotokat kijelentések halmazainak tekinthetjük. A logikák szabályai olyanok, hogy egy bemeneti állapotot, amely tartalmaz néhány bizonyos formájú állítást, adott formájú új állításokat származtathatunk, amelyeket hozzáadhatunk a kimeneti állapothoz.

Ez a folyamat, amelynek során logikai szabályokat alkalmazunk állításhalmazokra, hogy új állításokat hozzunk létre, a *dedukció* folyamata.

Bár ebben a fejezetben a dedukcióra fogunk koncentrálni, fontos tudatosítani, hogy léteznek másfajta érvelési módok is. A 15. fejezetben és a 19. fejezetben még hosszasan visszatérünk erre a kérdésre, de egyelőre csak röviden vizsgáljuk meg a dedukció és az *indukció* közötti különbséget.

Az indukció az empirikus tudomány által alkalmazott érvelési forma. Az induktív bizonyítás meglehetősen diffúz dolog a deduktív bizonyítástól. A deduktív bizonyítás annak bizonyítása - olyasféle, mint amelyet a következő részben látni fogunk -, hogy bizonyos állításformák más állításformákból bizonyos logikai szabályok szerint levezethetők. Egy tipikus *induktív* bizonyítás ezzel szemben bizonyos feltételek mellett végzett megfigyelések összegyűjtése annak az állításnak az alátámasztására, hogy a jövőben ilyen feltételek mellett ugyanezek a megfigyelések fognak történni.

A deduktív bizonyítások sajátossága, hogy nem vizsgálhatók felül. Egy deduktív bizonyításhoz semmi olyat nem lehet hozzáadni, ami a kiegészítés fényében nem vezetne az eredeti következtetéshez. Az induktív bizonyítás ezzel szemben alapvetően felülvizsgálható. Az induktív bizonyítások csak bizonyos valószínűséggel állapítják meg következtetéseiket, és mindig csak egy ellenérv van attól, hogy a feltételezett következtetésük nem teljesül.

Bár itt csak egy durva vázlatot adtam az induktív érvelésről, a deduktív érveléstől való megkülönböztetést két okból szemléltetem: először is azért, hogy megértse, hogy vannak olyan legitim és bevett érvelési módok, amelyek különböznek a deduktív érveléstől; másodsor azért, hogy felismerje, hogy a "bizonyíték" szó a tudósok szájából egészen mást jelent, mint a logikusok vagy matematikusok szájából. Amikor azt halljuk, hogy a tudósok "bebizonyítottak" valamit, vagy hogy "a tanulmányok kimutattak" valamit, érdemes tudatosítani, hogy a bizonyított vagy kimutatott dolgot a későbbi vizsgálatok fényében felülvizsgálhatják és megcáfolhatják.

A tudományos érvelésről még sok mindent el lehet mondani, és

biztosan nem szeretném, ha azzal vádolnának, hogy csak egy karikatúrát adtam elő

a tudományos folyamatról, ezért még egyszer utalok a további olvasásra vonatkozó javaslatokra.

Mielőtt a szakértői rendszerek vizsgálatát folytatnánk, szükségünk lesz egy kis terminológiára a *feltételes* és *predikátumokkal* kapcsolatban.

13.2 FELTÉTELESSÉG ÉS PREDIKÁCIÓ

A természetes nyelvi feltételes kifejezések a "ha akkor" formájú kijelentések. A feltételes kijelentések tanulmányozása és megfelelő formális leírásuk meghatározása központi jelentőségű a logikában. Számos logika kizárólag a feltételes feltételek kezelése alapján különböztethető meg.

A feltételes függvényeket egy nyíl segítségével ábrázolhatjuk. A 'ha ma hétfő van, akkor holnap kedd' állítás a következőképpen ábrázolható:

ma hétfő, holnap kedd

A feltételes feltétel bal oldala - amely a "ha" és az "akkor" közötti részt jelenti - a feltételes feltétel előzménye. A jobb oldali rész - az a bit, amely a 'then' után következik - a feltételes feltétel *következménye*.

Ha egy feltételes feltétel előfeltétele *teljesül*, akkor egy egyszerű logikai elv szerint levezethetjük a következményt. Ha tehát valóban igaz, hogy ma hétfő van, akkor - a fenti feltételes feltételezésből *kiindulva* - *levezethetjük*, hogy holnap kedd van. Ezt a logikai elvet *modus ponensnek* nevezik, és a következőképpen szimbolizálható:

<I 'f

<I

'f

A *modus ponens* logikai elve - amely azt mondja, hogy egy teljesített előfeltételű feltételből következtethetünk a következményre - az egyetlen logikai elv, amelyre a szakértői rendszerek vizsgálata során hivatkozni fogunk.

Mielőtt megnéznénk egy példa szakértői rendszert, utoljára a *predikátumokat* és a logikai formákat kell megvitatnunk.

Vegyük fontolóra a következő két állítást. Ha valami kutya, akkor emlős. Ha valami emlős, akkor van szíve. Ezeket az állításokat a következőképpen is ábrázolhatjuk:

valami egy kutya, ami egy emlős valami egy
emlős, aminek szíve van.

Ennél azonban jobbat is tudunk. Vegyük észre, hogy minden egyes elő- és utótagban egy tulajdonságot alkalmazunk egy dologra, vagy egy dologról *állítunk fel valamit*. Figyeljük meg azt is, hogy minden feltételes módban *ugyanaz a dolog*, amelyre mind az antecedens, mind a konzekvencia utal.

Ha a 'kutya' a kutya tulajdonságát, az 'emlős' az emlős tulajdonságát, a 'szív' pedig a szív tulajdonságát képviseli, akkor a fenti feltételeseket átfogalmazhatjuk úgy, hogy rögzítsük azt a tényt, hogy az elő- és utótagok tulajdonságait ugyanarra a dologra alkalmazzák:

kutya (x) emlős (x) emlős (x)
szív (x)

Ezek a feltételes alakok olyan közel állnak a logikai formához, amennyire e fejezet céljaira szükségünk van. A "kutya", az "emlős" és a "szív" szimbólumok - amelyek természetesen egységesen helyettesíthetők bármely más, általunk választott szimbólummal - *predikátumokat* jelölnek. Céljaink érdekében a predikátumokat úgy értelmezhetjük, mint amelyek tulajdonságokat és relációkat kódolnak.

Az x szimbólum a fenti feltételesekben egy változó - ahogyan azt kétségtelenül észrevetted. Azt fogjuk mondani, hogy az egyik feltételes feltétel *teljesül*, ha van egy olyan állításunk, amelynek ugyanez a *logikai formája*.

A mi céljaink érdekében a kijelentések predikátumokat alkalmaznak a *nevekre* (nem a változókra). Tehát ha például tudjuk, hogy Mia egy kutya, akkor ezt úgy tudjuk ábrázolni, hogy az m szimbólumot használjuk Mia nevéként, a következőképpen:

kutya (m).

Most a két feltételes függvényt, amelyet szimbolikusan reprezentáltunk, felhasználhatjuk néhány egyszerű következtetés elvégzéséhez. Az állítás - kutya (m) - ugyanolyan logikai formájú, mint az első feltételes feltételünk előfeltétele. Ez azt jelenti, hogy a feltételes feltétel előfeltétele *teljesül*, így *levezethetjük* a következményt, nevezetesen az emlős (m) állítást. Most már van egy olyan állításunk, amely kielégíti a második feltételes feltétel előfeltételét, így levezethetjük a következményét, és levezethetjük a szív (m)-et. Mivel tudjuk, hogy a "szív" a szívvel való rendelkezést jelenti, és hogy az m a Mia neve, az imént levezettük, hogy Mia-nak szíve van.

Ezt a következtetést szimbolikusan a következőképpen ábrázolhatjuk:

kutya (m)
kutya (x) emlős (x)

emlős (*m*)

emlős (x) szív (x)

szív (m)

13.1. gyakorlat

- (a) Ha valami nő, akkor az a valami ember. Ha valami ember, akkor az a valami halandó. Jelképesen ábrázoljuk ezt.
- (b) Sue egy nő. Mutassa be szimbolikusan azt a következtetést, amelyet ebből levonhat, az a) pontban bemutatott feltételes feltételek ismeretében.

Az eddig látott predikátumok *egyhelyű* vagy *monádikus* predikátumok voltak. Példánk szakértői rendszerében olyan predikátumokat is használni szeretnénk, amelyek két név közötti kapcsolatokat attribútumként adnak meg.

Tekintsük a "Mia idősebb, mint Linus" állításban kifejezett összefüggést. Az "idősebb, mint" reláció két olyan dolog között áll fenn, amelyeket *m-nek* és *l-nek* nevezhetünk, lehetővé téve a következő szimbolikus ábrázolást:

older_than (m , l)

Vegyük észre, hogy a nevek sorrendje a kéthelyű predikátumokban fontos. A következő állítás úgy értelmezhető, hogy Linus idősebb, mint Mia.

older_than (l , m)

Az eddig kidolgozott szimbolikát használhatjuk arra, hogy kódoljunk valamit, amit - a józan ész alapján - a "régebbi, mint" relációról tudunk, a következőképpen:

older_than (x , y) & older_than (y , z) older_than (x , z)

A fentiekben az amperjel (&) egyszerűen az "és" logikai műveletét, a *konjunkciót* jelöli. Azt mondjuk, hogy a fenti feltételnek *konjunktív* előzménye van. Az x , y és z szimbólumok változók, nem nevek.

A fenti feltételes mód azt mondja, hogy ha x idősebb, mint y , és y idősebb, mint z , akkor x idősebb, mint z . Ezt mindenki, aki érti az "idősebb, mint" jelentéseit, implicit módon megérti. Logikai értelemben ez azt jelenti, hogy az 'older_than' szimbólum egy *tranzitív* relációt jelöl.

13.2. gyakorlat

- (a) Milyen más relációkat tudsz még mondani, amelyek tranzitívak?
- (b) A *szimmetrikus* kapcsolat olyan, hogy ha x viszonyul y -hoz, akkor y is viszonyul x -hez. Milyen szimmetrikus kapcsolatokra tudsz gondolni?
- (c) Használja az általunk kifejlesztett szimbolikát annak ábrázolására, hogy egy adott összefüggés szimmetrikus.

Ha tudjuk, hogy Mia idősebb, mint Linus, és azt is tudjuk, hogy Sue idősebb, mint Mia, akkor az s - t Sue névként használva a következőképpen érvelhetünk:

older_than (s , m)
 older_than (m , l)
 older_than (x , y) & older_than (y , z) older_than (x , z)

older_than (s , l)

A feltételes feltétel konjunktív előfeltétele teljesül, mivel mindkét *konjunktív* - az amperjelet szegélyező állítások - teljesülnek, így engedélyünk van arra, hogy a konklúziót *modus ponens* módon levezessük, mint korábban. Biztos vagyok benne, hogy önök eléggé képesek felismerni, hogy mi az, amit bizonyítottunk.

Ez az összes terminológia és szimbolika, amire szükségünk van egy szakértői rendszer kifejlesztéséhez.

13.3 KINSHIP

Ha az előző fejezetben leírt példa- levezetéseket követni tudta, akkor már megértette a szakértői rendszerek működésének fontos aspektusait. Valójában a predikátum-jelölés bevezetéséhez használt példaesetek valójában maguk is miniatűr szakértői rendszerek voltak. A szakértői rendszerek olyan formális rendszerek, amelyek célja, hogy kódolják azokat az információkat, amelyeket egy releváns emberi szakértő egy adott tudásterületről tud, és hogy reprodukálják a deduktív eljárásaikat ezen információk és néhány új bemeneti adat birtokában. A mi példánk szakértő a rendszer a rokonsági kapcsolatokra vonatkozó információkat fogja kódolni.

A szakértői rendszer *rezidens információi* számos feltételes feltétel formájában kerülnek meghatározásra. Ez a rezidens információ szolgál a szakértői rendszer szabályaiként.

Míg mi egy logikai elvre - *modus ponens* - fogunk hivatkozni -

a szabályok *alkalmazása* során a következtetések levonásához, maguk a szabályok nem logikai szabályok. A rendszer szabályai inkább feltételes szabályok, amelyek azokat az információkat reprezentálják, amelyeket a szakértő a releváns tudástartományról - ebben az esetben a rokonságról - tud. A logikai elvet a szabályok alkalmazásának effektív eljárásában fogalmazzuk meg.

A rokonsági rendszerünk lakosainak adatait a következőképpen határozzuk meg:

grandparent_of (x , y) & male (x) grandfather (x) parent
 (x) & male (x) father (x)
 szülő (x) & nő (x) anya (x) parent_of (x
 , y) parent_of (x) parent_of (x , y)
 child_of (y , x)
 parent_of (x , y) & parent_of (y , z) grandparent_of (x , z)

A szakértői rendszer állapotai állítások halmazai. A rokonsági rendszerünk kezdeti állapota a következő lesz:

parent_of (j , m)
 parent_of (m , h)
 male (j)
 férfi (h) nő
 (m)

Az utolsó dolog, amit rokonsági rendszerünkhöz meg kell határozni, az az effektív eljárás, amellyel a szabályokat az államokra alkalmazzuk.

1. A rezidens információ első feltételesétől kezdve ellenőrizze, hogy van-e olyan állítás az állapotban, amely kielégíti az előfeltételt - azaz ellenőrizze, hogy az állapot bármelyik állítása ugyanolyan logikai formájú-e, mint az előfeltétel, és csak a változó(k) név(k) helyettesítésében tér el tőle.
2. Ha van olyan állítás, amely kielégíti az előfeltételt, akkor adjuk hozzá a következményt a levezetett állítások listájához (ügyelve arra, hogy a következményben szereplő változó(ka)t ugyanazzal a névvel helyettesítsük, mint az előfeltételt kielégítő állítás(ok)ban). Ellenőrizzük, hogy vannak-e további olyan állítások az állapotban, amelyek kielégítik az előfeltételt.
3. Ismétlje meg az 1. és 2. lépést a rezidensek adatainak minden egyes feltételére. Ha ez megtörtént, egészítse ki az eredeti állapotot a levezetett állítások listájával, és adja ki ezt a kiegészített állapotot. Csak akkor adjunk hozzá egy állítást a levezetett listából, ha az még nem szerepel az állapotban.
4. Kezdjük újra az első feltételes feltételes tételt, és nézzük meg, hogy a levont új állítások lehetővé teszik-e további új állítások levezetését.

5. Ha egy állapot olyan, hogy a rezidens információk szereplő feltételes feltételek egyike sem teszi lehetővé olyan állítások levezetését, amelyek még nem szerepelnek az állapotban, akkor állj.

Alkalmazzuk ezt az eljárást a kezdeti állapotunkra, és nézzük meg, mit tudunk levezetni.

13.3. gyakorlat

Mielőtt tovább olvasnánk, próbáljuk meg az eljárás első három lépését alkalmazni a rokonsági rendszerünk kezdeti állapotára.

Az első feltételes feltétel a rezidens információinkban

a következő: $\text{grandparent_of}(x, y) \ \& \ \text{male}(x)$

$\text{grandfather}(x)$

A kezdeti állapotunkban nincsenek $\text{grandparent_of}(x, y)$ típusú utasításaink, így nem tudjuk teljesíteni a feltételes feltétel előfeltételét. Az egyik kötőszót kielégíthetjük, mivel van $\text{male}(x)$ alakú állapotunk, de *mindkét* kötőszót ki kell elégítenünk ahhoz, hogy a kötőszó előfeltételét kielégítsük.

A következő két feltételes feltétel a rezidens információinkban:

$\text{szülő}(x) \ \& \ \text{férfi}(x)$ $\text{apa}(x)$ $\text{szülő}(x) \ \& \ \text{nő}(x)$ $\text{anya}(x)$

Ismétlem, nincs $\text{parent}(x)$ alakú állításunk, így nem tudjuk teljesíteni egyik feltételes feltétel előfeltételét sem.

Legyünk itt óvatosak, mert bár tudjuk, hogy $\text{parent_of}(j, m)$, és tudjuk - a józan ész alapján -, hogy ha valaki valaki másnak a szülője, akkor *ipso facto* szülő, nem feltételezhetjük csak úgy, hogy $\text{parent}(j)$. A parent és a parent_of szimbólumok csak szimbólumok. Olyan szimbólumokat használtunk, amelyek számunkra értelmesek, de ami a rendszer műveleteit illeti, a szimbólumok által reprezentált predikátumok közötti kapcsolatokat explicit módon kell kódolni a szabályokban. A következő feltételes feltétel pontosan ezt teszi.

A negyedik feltétel a rezidens információinkban a

következő: $\text{parent_of}(x, y)$ $\text{parent}(x)$

Van egy olyan állításunk a kezdeti állapotban, amely kielégíti ennek a feltételnek az előfeltételét - $\text{parent_of}(j, m)$ -, így hozzáadhatjuk a következményt (ügyelve arra, hogy a változó helyes nevével helyettesítsük) a levezetett állítások listájához - nevezetesen $\text{parent}(j)$.

Van egy másik állításunk is a kezdeti állapotban, amely kielégíti ugyanennek a feltételnek az előfeltételét - $\text{parent_of}(m, h)$ -, így a $\text{parent}(m)$ állítást is hozzáadhatjuk a levezetett állítások listájához.

Az ötödik feltétel a rezidens információinkban a

$\text{parent_of}(x, y)$ $\text{child_of}(y, x)$

A kezdeti állapotban két olyan állításunk van, amely kielégíti ennek a feltételnek az előfeltételét - $\text{parent_of}(j, m)$ és $\text{parent_of}(m, h)$ -, így hozzáadhatjuk a $\text{child_of}(m, j)$ és $\text{child_of}(h, m)$ állításokat a levezetett állítások listájához.

A végső feltételes feltétel a rezidens információinkban a

következő: $\text{parent_of}(x, y)$ & $\text{parent_of}(y, z)$

$\text{grandparent_of}(x, z)$

Itt óvatosnak kell lennünk - a névnek, amelyet az előtag első kötőszavában az y helyébe teszünk, ugyanannak a névnek kell lennie, amelyet a második kötőszavában az y helyébe teszünk. Mint kiderül, vannak olyan állapot- mensek, amelyek kielégítik ennek a feltételes feltételnek az előfeltételét - $\text{parent_of}(j, m)$ és $\text{parent_of}(m, h)$ -, ami lehetővé teszi számunkra a $\text{grandparent_of}(j, h)$ levezetését.

Most már minden egyes feltételes állítást megvizsgáltunk a rezidens információban, és öt új állítást vontunk le:

$\text{parent}(j)$

$\text{parent}(m)$

$\text{child_of}(m, j)$

$\text{child_of}(h, m)$

$\text{grandparent_of}(j, h)$

Tehát hozzáadjuk a levezetett utasítások listáját a kezdeti állapotunkhoz (ellenőrizve, hogy egyik sem redundáns), hogy a következő állapotot kapjuk:

$\text{parent_of}(j, m)$

$\text{parent_of}(m, h)$

$\text{male}(j)$

férfi (h) nő

(m) szülő (j)

szülő (m)

gyermek_a (m

, j)

$\text{child_of}(h, m)$

$\text{grandparent_of}(j, h)$

A következő teendőnk az, hogy sorban újra ellenőrizzük az egyes feltételeseket, hogy a származtatott állapotunkban lévő új állítások lehetővé teszik-e további új állítások levezetését.

13.4. gyakorlat

Mielőtt tovább olvasnánk, határozzuk meg, hogy a rezidens információkban szereplő feltételes feltételeken való második átfutás során milyen új állítások vezethetők le, ha vannak ilyenek. Mi a helyzet a harmadik menetben?

A rezidens információkban szereplő feltételes feltételek második átfutása lehetővé teszi számunkra, hogy három újszerű állítást következtessünk:

nagyapa (j) apa
(j) anya (m)

A harmadik menet csak felesleges utasításokat generál - olyanokat, amelyek már szerepelnek a generált állapotunkban -, így megállunk.

13.5. gyakorlat (Kihívás)

Kiegészítsük a rokonsági rendszerünk kezdeti állapotát a kijelentésekkel:

parent_of (m, j)
female (j)
diff (j, h)
diff (m, h)

és a következő feltételekkel egészítse ki a rezidensek adatait:

parent_of (x, y) & parent_of (x, z) & diff (y, z) siblings
(y, z)
testvérek (x, y) & férfi (x) testvér (x)
testvérek (x, y) & nő (x) nővér (x)
testvérek (x, y) testvérek (y, x)
testvérek (y, x)

Alkalmazza a szabályokat az összes lehetséges kimutatás létrehozásához. Kicsit furcsának tűnnek a dolgok? Miért?

13.4 EXPERT RENDSZEREK

Az előző szakasz szakértői rendszere jelentősen leegyszerűsített, de ennek ellenére egyértelmű, hogy képes megragadni azt a deduktív folyamatot, amelyet a rokonsági kapcsolatokról való gondolkodás során végzünk.

Ez nehézkesnek és mesterségesnek tűnhet a gondolkodási folyamatainkhoz képest, de ez csak azért van, mert az a fajta érvelés, amit akkor végzünk, amikor azt mondjuk, hogy valaki nő szülő, gyors és automatikus. Nincs szükségünk arra, hogy *kifejezetten* alkalmazzunk egy szabályt ahhoz, hogy megállapítsuk, hogy a kérdéses személy anyja - ezt egyszerűen automatikusan megértjük, amikor megértjük, hogy az illető női szülő. Ez azonban nem szól ellene annak az állításnak, hogy ezt az implicit megértést éppen ilyen módszerek irányítják.

Valójában, ha arra kéri Önt, hogy határozza meg, milyen kapcsolatban áll Önnel édesanyja húgának lánya lányának férje, akkor valószínűleg sokkal kifejezettebben kell gondolkodnia az érintett rokonsági kapcsolatokról, nagyon hasonló szabályokat követve, mint amilyeneket az előző szakaszban kódoltunk.

Az érdekes szakértői rendszerek természetesen jóval összetettebbek, mint a mi példánk. A mi rokonsági rendszerünk nagyon kevés rezidens információval rendelkezik, és csak egyetlen logikai elvre - *modus ponens* - apellál. A bonyolultabb szakértői rendszerek jóval több rezidens információt tartalmaznak, és a szabályok alkalmazása során számos logikai elvre hivatkoznak.

Következésképpen gyakran szeretnénk egy alulról felfelé irányuló keresést generálni annak megállapítására, hogy egy adott állítás szerepel-e bármely generált állapotban. Ilyen esetekben csak azzal az állítással kezdünk, amelynek a levezetésében érdekeltek vagyunk, és visszafelé haladunk a szabályokon keresztül, nem az egész állapotokat, hanem csak azokat az állításokat vesszük figyelembe, amelyeknek szerepelniük kell egy állapotban ahhoz, hogy az előző iterációban levezessük az állítás(oka)t. Ha egy csomópontban lévő állítás szerepel a kezdeti állapotban, akkor azt a következő iterációnál kihúzhatjuk. Ha olyan csomóponthoz jutunk, ahol az előző iterációban levezetett állítás bizonyításához semmi sem kell, hogy szerepeljen az állapotban - azaz az állítás(ok), amelyre szükségünk volt a minket érdeklő állapot generálásához, bizonyítottan szerepel(nek) a kezdeti állapotban -, akkor a keresés sikerrel zárul.

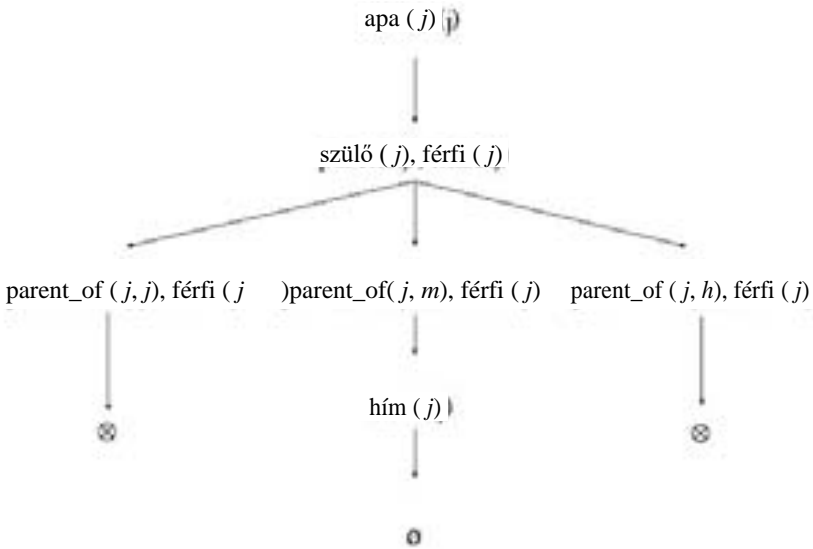
A 13.1. ábra egy alulról felfelé irányuló keresést mutat be az állam apja számára.

(j) a rokonsági rendszerünkben. Az első iterációnál visszafelé haladva a rendelkezésre álló szabályokon keresztül megállapítjuk, hogy apa (*j*) csak akkor generálható, ha van szülő (*j*) és férfi (*j*).

A második iterációnál három ág van, amelyek azt a három utasítást képviselik, amelyek lehetővé teszik számunkra, hogy a (*j*) szülőt létrehozzuk. A bal és a jobb oldali ágak kudarcra végződnek, mivel sem a *parent_of (j, j)*, sem a *parent_of (j, h)* *nincs* a kezdeti állapotban, és nincs olyan szabály, amely lehetővé tenné a *parent_of*

utasítások generálását.

A középső ág folytatódik, mivel a $\text{parent_of}(j, m)$ a kezdeti állapotban van, így ezt az állítást a következő leszálló csomópontban feloldjuk, így csak a $\text{male}(j)$ marad bizonyításra. A következő iterációnál feloldjuk a



13.1 ábra Alulról felfelé irányuló keresés.

ez utóbbi állítás, mivel ez is a kezdeti állapotunkban van, így nem marad semmi (az üres halmaz), amit bizonyíthatnánk. Tehát a középső ágon végigvezető keresési út sikerrel végződik, és az ágot követve visszaolvassuk az off levezetést a gyökércsomópontig.

13.6. gyakorlat

Generáljunk alulról felfelé irányuló keresést a rokonsági rendszerünkben az államokra:

- (a) anya (m);
- (b) nagyapa (j).

A szakértői rendszerek hasznosan alkalmazhatók a gyenge mesterséges intelligencia projektekben. Az egyik ilyen jól ismert rendszer a MYCIN, amelyet az 1970-es években fejlesztettek ki a Stanfordon. A MYCIN képes valószínűségi diagnózisokat felállítani a betegségekről, és vérvizsgálatok eredményei alapján gyógyszert ajánlani. A rezidens információi heurisztikus eljárásokat kódolnak, amelyeket az orvosok követnek, amikor fejlett kultúrák hiányában durva és kész diagnózisokat állítanak fel. Szabályalkalmazási eljárása a valószínűségi gondolkodás elveire hivatkozik, így képes számos lehetséges diagnózist javasolni, mindegyikhez egy-egy bizonyossági tényezőt rendelve.

A mi céljaink szempontjából még érdekesebb, hogy vannak olyanok, akik reménykednek abban, hogy a szakértői rendszerek önmagukban erős mesterséges

intelligencia. A Douglas Lenat által alapított és a Cycorp magán kutatóintézet égisze alatt fejlesztett Cyc projekt pontosan ezt a célt szolgálja. A Cyc kutatóinak meggyőződése az, hogy ha a rendszer rezidens információiban kódolni tudják az összes (vagy sok) olyan információt, amit önök és én általános tudásnak tekintünk, és egy megfelelően kifinomult következtető motort fejlesztenek ki a következtetések levonásához, akkor egy erős mesterséges intelligencia-artefaktumot fejlesztenek ki.

Az, hogy a szakértői rendszerek módszerei alkalmasak-e a mesterséges intelligencia számára, továbbra is nyitott empirikus kérdés. Van azonban okunk azt hinni, hogy e megközelítéssel problémák adódhatnak. Ezek közül néhányat a 15. fejezetben, néhányat pedig a 19. fejezetben fogunk megvizsgálni. A Cyc projekt védelmében azonban úgy tűnik - abból, amit a működésükről megtudtam -, hogy legalább néhány ilyen aggályra reagálnak, és ennek megfelelően bővítik a hagyományos szakértői rendszermodellt.

Talán a legjelentősebb és legnehezebb számítási probléma, amelyet a valódi mesterséges intelligencia felé vezető úton meg kell oldani, a természetes nyelv értelmezésének és előállításának problémája. A következő fejezetben erre a kérdésre fordítjuk figyelmünket.

GÉPEK ÉS NYELV

A természetes nyelv kezelésére szolgáló számítási eljárások kidolgozása vitathatatlanul a mesterséges intelligencia kutatóinak legjelentősebb problémája. Ebben a fejezetben azzal kezdjük, hogy megvizsgáljuk azokat a különböző számítási problémákat, amelyeket meg kell oldani ahhoz, hogy ezt megkönnyítsük.

Ahogy a gépi következtetés esetében, a nyelvi értelmezést és produkciót megkönnyítő számítógépes módszerek átfogó áttekintése külön köteteket igényelne. Itt most csak egy ilyen módszerrel foglalkozunk, amely a nyelvi kompetencia egyik konstitutív funkciójának - a nyelvhelyesség meghatározásának - szolgálatában áll.

A 16. fejezetben tovább vizsgáljuk a nyelvi tevékenységet irányító eljárásokat, és a 19. fejezetben visszatérünk a nyelvi képességet kiszolgáló további funkciókat megvalósító számítógépes módszerek vizsgálatára.

14.1 TOLMÁCSNYELV

Gondolkodjunk el a különböző eljárásokon, amelyek a beszélt beszéd megértésében szerepet játszanak.

A beszélt nyelvet általában folyamatos hangfolyamban adják elő, amely nem mutatja ki könnyen nyelvi tulajdonságait. Bár úgy tűnhet, hogy az egyes szavak könnyen megkülönböztethetők pusztán a hangtani tulajdonságok alapján, ez általában nem így van, hacsak nem beszélünk - nagyon - lassan - és - óvatosan. Ez világosan látszik, ha megnézzük egy rögzített beszéd hullámformájának vizualizációját. Továbbá, a kimondás *abszolút* hangtani tulajdonságai - mint például a hangmagasság és a hangerő - beszélőnként jelentősen eltérnek egymástól. A beszélt beszéd értelmezése nem triviális feladat - valójában meglehetősen összetett eljárás.

Számos olyan feladat van, amely a hallott input természetes nyelvi mondatként való értelmezését közvetíti. Az egyik ilyen feladat

a fonetikai bemenet *fonémikus* reprezentációvá történő átalakítása.

A fonémák azok az atomi értelemmel bíró beszédhangok, amelyekből a beszélt nyelv felépül. A 16. fejezetben sokkal többet fogunk megtudni a fonémákról, de egyelőre egy durva leírás is elegendő.

A fonémák olyan eszményképek, amelyekhez a tényleges beszédhangok - a *telefonok* - közelítenek, és amelyek egy adott nyelvben *jellegzetes kontrasztokat* képviselnek. A "pat" szó kezdőhangja és a "bat" szó kezdőhangja két olyan fonéma, amelyek csak a *hangalak tekintetében* különböznek. Az /p/ nem hangzó, de a /b/ igen. Ez azt jelenti, hogy a /b/ fonéma kiejtésekor a hangszalagok rezegnek, de a /p/ hang kiejtésekor a levegő csak hangtalanul áramlik át. Próbálja meg az ujjait a hangszalagjára helyezni, miközben a "pat" és a "bat" szót ejti, és meg fogja tudni állapítani a hangképzés közötti különbséget. Abból a tényből, hogy a 'pat' és a 'bat' differenciális dolgokat *jelent*, megállapíthatjuk, hogy a hangképzés egy *megkülönböztető ellentét* az angolban, és hogy az /p/ és a /b/ különálló fonémák.

Ez a fonémák nagyon durva és kész leírása, de jelen célokra csak azt kell megértenünk, hogy a beszélt beszéd értelmezésének egyik közvetítő feladata az, hogy a beszéd fonetikai tulajdonságaiból meghatározzuk a kimondott nyelvben értelmezhető, egymástól elkülönülő, kontrasztos hangegységek reprezentációját.

Egy másik közvetítő feladat a fonémikus ~~paradigma~~ szintaktikai struktúrába való bontása. Más szóval, meg kell határoznunk, hogy az értelmes beszédhangok sora hogyan bomlik szavakra, mondatokra, mondatrészekre és mondatokra. Végül ki kell dolgoznunk, hogy ezek a szavak, mondatok és mondatok mit *jelentenek*.

Összességében tehát három különböző reprezentációs szint van jelen egy kifejezés megértésében: a fonémikus, a szintaktikai és a *szemantikai* reprezentációk.

Egy kijelentés előállítása ugyanezt a három reprezentációs transzformációt foglalja magában. Meghatározunk egy bizonyos jelentést, meghatározzuk a jelentést kódoló szintaktikai struktúrát, átalakítjuk a szintaktikai reprezentációt fonémás reprezentációvá, és végül minden egyes fonéma esetében fonetikus kimenetet állítunk elő. A fonetikai kimenet előállítását az a fonémakörnyezet irányítja, amelyben a fonéma található - a nyelvi produkció egy olyan jellemzője, amelyet a 16. fejezetben fogunk megvizsgálni.

Bár az a szekvenciális mód, ahogyan ezeket az átalakulási szakaszokat felsoroltam, arra engedhet következtetni, hogy a feldolgozás egyes szakaszai külön-külön és egymás után következnek be, jó bizonyítékok szólnak amellett, hogy

mindhárom szakasz együttesen zajlik.

Úgy tűnik például, hogy a szintaktikai struktúra meghatározását befolyásolják azok az elvárások, amelyeket az a szemantikai reprezentáció irányít, amelyet a kifejezés szintaktikai elemzése során építünk fel. Hogy ezt világossá tegyük, tekintsük a következő három mondatot.

- [1] A ló elszáguldott az istálló mellett.
- [2] A ló tegnap elszáguldott az istálló mellett.
- [3] A ló elszáguldott az istálló mellett, és elesett.

Az [1] mondat megfelel az angol *nom* prototipikus szintaktikai mintájának - egy főnévi igenévből áll, amelyet közvetlenül egy igekötő követ, amely egy tranzitív igéből és annak komplementjéből áll. A [2] mondat szintén megfelel ennek a mintának - csupán egy toldalékot ad az igekötő után.

A [3] mondat viszont kezdetben szokatlannak és potenciálisan nyelvtanilag nem megfelelőnek tűnik. Hallunk egy főnévi igenevet "a ló", amelyet azonnal követ a mondat fő igéjének vélt "versenyzett", és azt várjuk, hogy a mondat (vagy legalábbis a mondat) többi része az ige kiegészítéséből (az istálló mellett / három órákor Flemingtonban) és esetleg egy vagy több toldalékból (meglehetősen gyorsan / múlt kedden) álljon.

Következésképpen, amikor a "past the barn" után az "esett" ige következik, megdöbbenünk, mivel talán egy toldalékra számítottunk, de nem egy másik igére. Eddig a pontig már meghatároztuk a mondat szemantikai reprezentációját, de az utolsó ige jelenléte azt mutatja, hogy vagy tévedtünk a jelentéssel kapcsolatban (amikor azt hittük, hogy ez az [1] jelentése), vagy a mondat nem grammatikus.

Más szóval, a jelentés, amelyet a mondat értelmezésekor konstruálunk - azzal együtt, amit implicit módon tudunk a tipikus angol mondatok szintaktikájáról - meghatározza a ~~szintaktikai~~ szerkezetre vonatkozó elvárásainkat. Feltételezzük a legvalószínűbb szintaktikai struktúrát, és ennek megfelelően strukturáljuk szemantikai reprezentációinkat, ami aztán befolyásolja a fennmaradó szintaktikai elvárásainkat.

Valójában elég nehéz lehet túllépni a [3] látszólagos ungrammatikusságán, amíg rá nem jövünk, hogy a mondat nem az eredetileg feltételezett tipikus szintaktikai struktúrával rendelkezik. A 'esett' ige a [3] főigéje, a 'vágatott az istálló mellett' pedig csupán azt jelzi, *hogy melyik* ló esett el - az, amelyik vágatott az istálló mellett. Megjegyezzük, hogy a 'raced past the barn' helyettesíthetjük a 'painted purple stripes' kifejezéssel, hogy hasonló effektust érjünk el, bár az effektus gyengébb lesz, mivel kevésbé valószínű, hogy a lóról feltételezzük, hogy a festő ágens, mint a versenyző ágens. Ha azonban a "versenyzett az istálló mellett" helyett a "gyorsan futott" vagy az "Anne-hez tartozott" kifejezést használjuk, a mondat

egyértelművé válik.

A kétértelmű szintaktikájú mondatok bizonyítékot szolgáltatnak arra, hogy a szemantikai reprezentációk és a szintaktikai reprezentációk együttesen és egymást kölcsönösen befolyásolva épülnek fel. A homofonikus mondatok további bizonyítékot szolgáltatnak arra, hogy a fonémikus reprezentációk szintaktikai elemekre való bontását mind a szintaktikai, mind a szemantikai elvárások befolyásolják.

A homofonok olyan fonémák sorozatai, amelyeknek egynél több szemantikai értelmezése van, mint például a "pár" és a "körte". A homofónia speciális esete a homonímia. A homonimák olyan homofonok, amelyeknek ortográfiai megjelenítése is azonos, azaz ugyanúgy írják őket, mint például a 'bank' ige és a 'bank' főnév. Az angol nyelvben nagyon sok homonímia van, mivel sok szó értelmezhető főnévként vagy igeként - pl. paint, spring, void, power, urge, whisper, sleep stb.

A homofonikus mondatok olyan fonémák sorozatai, amelyek egynél több különböző szintaktikai szerkezetként elemezhetők. Egy példa homofonikus frázisra a fonémák olyan sorozata, amely a következőképpen értelmezhető: "magasan fent" vagy "mérj egy pitét".

Értelemzhetjük az *Óz, a nagy varázslóban* Judy Garland által énekelt sort.

a következő két mondat valamelyikeként:

[4] Valahol, a szivárványon túl, nagyon magasan.

[5] Valahol, a szivárványon túl, mérj egy pitét.

Mindkettő tökéletesen nyelvtani szintaktikai szerkezet - a [4] egy leírás, az [5] pedig egy imperatívusz. Következésképpen a szintaktikai megfontolások önmagukban nem elegendők ebben az esetben a két lehetséges értelmezés közötti különbségtételhez.

A mi szemantikai értelmezésünk azonban erősen a [4] felé hajlik. A "Valahol, a szivárványon túl" szemantikai értelmezése sokkal valószínűbbé teszi azt a hipotézist, hogy Judy Garland a továbbiakban ezt a helyet írja le, és *ipso facto* sokkal kevésbé valószínű hipotézis, hogy egy pite mérlegelési parancsot ad ki.

Az olyan esetekben azonban, mint a következő, nem egyértelmű, hogy a homofonikus kifejezés disambiguációzását szintaxis vagy szemantikai megfontolások vezérlik:

[6] Elmentem a pékségbe, hogy megmérjem a pitét.

[7] * I went to the baker's to way up high

A [7] mondat nem grammatikus (amint azt a csillag jelzi), így ebben az esetben a szintaktika önmagában elegendő a homofon egyértelművé tételéhez; azonban diffikult a szintaktikai megfontolást a szemantikai asszociáció hatásától elválasztani.

Szerencsére ez nem olyan dolog, amiről itt döntenünk kell.

A fejezet hátralévő részében arra a tényre szeretnék koncentrálni, hogy a [7]-et könnyen és automatikusan felismerjük mint nem ~~gramm~~mondatot. Azt szeretném megvizsgálni, hogyan alkalmazhatnánk szimbólumrendszeres módszereket az írott nyelvi karakterláncok nyelvtani helyességéről való döntéshez.

14.2 GENERATÍV NYELVTAN

Noam Chomsky az 1950-es években forradalmasította a nyelvtudományt azzal, hogy új megközelítést alkalmazott a nyelvtan tanulmányozásában.

Chomsky előtt a nyelvtan alig tartalmazott többet, mint a beszédrészek taxonómiáját és a nyelvtant tanulók által betartandó elvek felsorolását. Például az előíró nyelvtanok tanulóinak azt mondták, hogy soha ne hasítsanak infinitívuszt.

Chomsky ezzel szemben egy nyelv nyelvtanát úgy tekintette, mint azt a mechanizmust, amellyel a nyelv összes és csakis a nyelv grammatikai karakterláncai *létrehozhatók*. Chomsky azt állította, hogy a nyelvünk szisztematikus *generatív* nyelvtanának internalizált ismerete magyarázza a nyelv végtelen *produktivitását*.

A nyelv végtelenül produktív, mivel végtelen számú mondatot vagyunk képesek létrehozni, és annak nyelvtani helyességéről tudunk dönteni, annak ellenére, hogy csak véges számú mondattal találkoztunk. Anyanyelvi beszélőként azonnal meg tudja mondani, hogy a "Michelle nagyanyja drogot árul motorosoknak Fehéroroszországban" mondat az angol nyelv nyelvtani mondata, még akkor is, ha nem valószínű, hogy valaha is találkozott már ezzel a bizonyos mondattal.

A formális rendszerek a nyelv produktivitását elősegítő mechanizmusok elsőszámú jelöltjei. Láttuk, hogy a formális rendszerek hogyan képesek véges erőforrásokból végtelen számú állapotot rekurzív módon, szabályvezérelt módon létrehozni. Láttuk azt is, hogyan végezhetünk alulról felfelé irányuló keresést annak megállapítására, hogy egy adott rendszerben egy adott állapotot generálnak-e vagy sem. Most azt fogjuk megnézni, hogyan használhatjuk a formális rendszereket egy nyelv *generatív* nyelvtanának meghatározására.

Egy nyelv generatív nyelvtana egy sajátos formális rendszer. Ez egy szimbólumrendszer, amely hasonló a 7. fejezetben ismertetett [STR] és [BIN] rendszerekhez. Egyes szimbólumai - amelyek a generációs fa terminális csomópontjaiban jelennek meg - a nyelv lexikai elemeiként (szavaként) értelmezhetők. A többi szimbólumot nyelvtani kategóriákként értelmezhetjük, mint például "mondat", "főnévi kifejezés", "melléknév" stb.

A generatív nyelvtan szabályai olyan átírási szabályok, mint az

[STR] és a [BIN] szabályai. Az [STR] rendszerben az újraírási szabályok a *kontextusban* voltak

függő - az, hogy alkalmazhatunk-e egy szabályt egy szimbólumra vagy sem, az állapot környező szimbólumaitól függ. A [BIN] rendszer azonban rendelkezett *kontextusmentes* újraírási szabályokkal.

A generatív nyelvtan kizárólag olyan kontextusmentes újraírási szabályokkal rendelkezik, amelyek úgy vannak kialakítva, hogy bármely szabály bemeneti oldalán csak egy szimbólum található. Egy olyan formális rendszert, amely megfelel ezeknek a kritériumoknak, *mondatszerkezeti nyelvtannak* nevezünk.

Adott egy mondatszerkezeti nyelvtan, a nyelvtan szerinti összes és csakis a nyelvtani karakterláncokat generálhatjuk *mondatszerkezeti fák* létrehozásával. A mondatszerkezeti fa olyan, mint az eddig látott generációs fák, egy kivétellel. Míg korábban a csomópontok állapotokat tartalmaztak, addig a mondatszerkezet-fa csomópontjai egyenként csak egyetlen szimbólumot képviselnek, a leszármazó csomópontjai pedig azt a szimbólumot vagy azokat a szimbólumokat, amelyekkel át van írva. A nyelvtan által megadott nyelvtani karakterláncokat a mondatszerkezet-fa terminális csomópontjain keresztül olvassuk be.

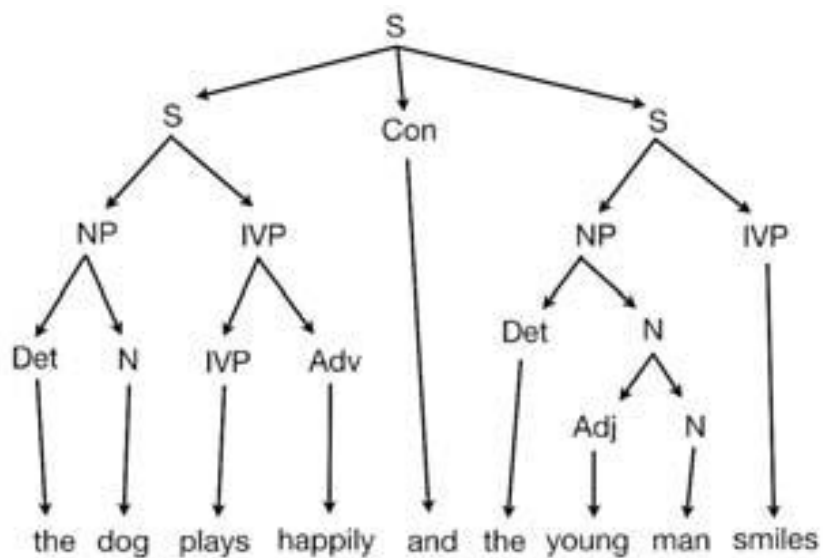
Állítsunk fel egy példát a mondatszerkezeti nyelvtanra, hogy mindez világosabbá váljon.

14.3 MONDATSZERKEZETI FÁK

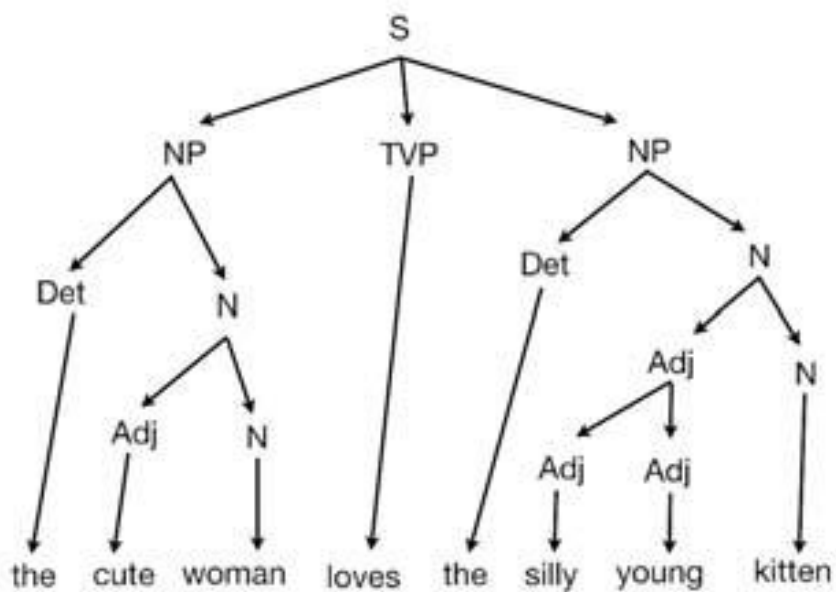
Megadunk egy mondatszerkezeti nyelvtant az angol nyelv egy töredékére. A mondatszerkezeti nyelvtanunk állapotai a szabályokban szereplő szimbólumok véges karakterláncai lesznek. A kezdeti állapot az 'S' szimbólum lesz. A rendszer szabályai a következők.

S	S Con S / NP IVP / NP TVP NP NP
Con	és / vagy / de
NP	Det N
Det	a / a
N	Adj N
N	férfi / nő / cica / kutya
Adj	Adj Adj
Adj	fiatal / boldog / aranyos / buta
IVP	IVP Adv
IVP	fut / eszik / játszik / mosolyog
Adv	gyorsan / szépen / boldogan
TVP	szeret / undorodik / akar
PP	P NP
P	a címre.

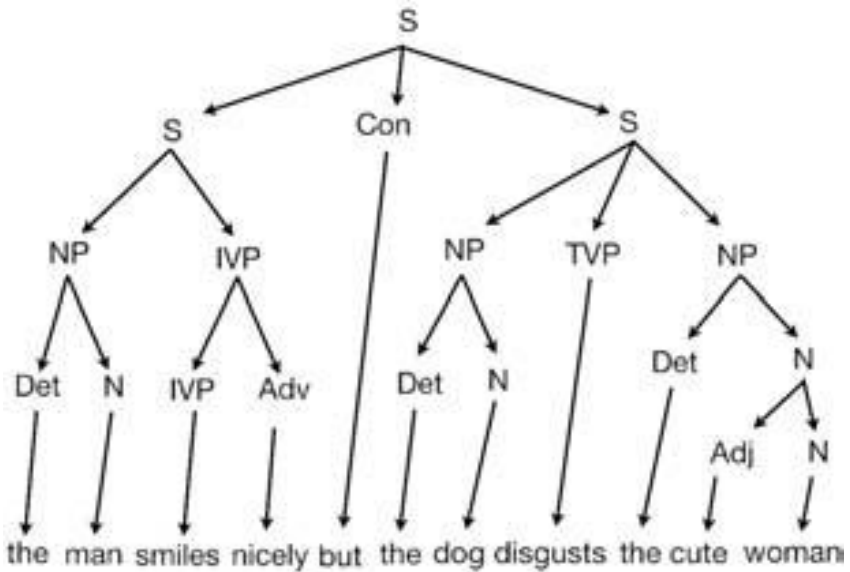
Ezt a mondatszerkezeti nyelvtant használhatjuk nyelvtani karakterláncok létrehozására, amelyeket a 14.1., 14.2. és 14.3. ábrán látható mondatszerkezeti fák terminális csomópontjai képviselnek.



14.1. ábra A mondatszerkezeti fa.



14.2. ábra A mondatszerkezeti fa.



14.3 ábra A mondszerkezet fája.

14.4 SZÁMÍTÁSTECHNIKAI NYELV

A nyelvi mondatok nyelvtani jellegének meghatározása egy generatív nyelvtan szerint egyértelműen számítási eljárás. Az is világosnak kell lennie, hogy adott egy mondszerkezeti nyelvtan és egy tetszőleges szimbólumsorozat, alulról felfelé irányuló keresést végezhetünk annak megállapítására, hogy a sortozatot egy mondszerkezeti fa generálja-e vagy sem.

Az előző felelített példamondszerkezeti nyelvtan természetesen erősen leegyszerűsített, és az angol nyelv lexikai elemeinek és szintaktikai struktúráinak csak egy kis töredékét veszi figyelembe. Egy természetes nyelv teljes generatív nyelvtanának megalkotásához nem csak azoknak a szabályoknak a megadása szükséges, amelyek alapján a mondszerkezeti fák felépítjük, hanem a mondszerkezeti fák terminális láncainak különböző szintaktikai átalakításai is, amelyek az anyanyelvi beszélők által előállított számtalan mondatért felelősek.

A szintaktikai átalakulás ezen további elemével itt nem kell foglalkoznunk. Mint mindig, az érdeklődő olvasó követheti a további olvasmányokra vonatkozó javaslatokat, vagy elvégezhet egy bevezető kurzust a generatív nyelvtanból. Céljainkhoz elegendő a következő megállapításokat tenni.

Úgy tűnik, hogy azok a mechanizmusok, amelyek megkönnyítik a nyelvtani ítéleteket az anyanyelvi beszélőkben, számítással is megvalósíthatók. Tekintettel arra, hogy ez az egyik olyan funkció,

158
amely a különféle

a természetes nyelvi kifejezések megértéséhez szükséges reprezentációs transzformációkat, némi előrelépést értünk el a nyelvi képességek számításhoz való leírása felé.

Most már világosnak kell lennie, hogy a lin- gvisztikus viselkedés milyen bonyolult. Bár az írott és beszélt nyelv előállítására és megértésére olyan természetes számunkra, hogy a legegyszerűbb folyamatnak tűnik, valójában számos közvetítő eljárás segíti a nyelvi tevékenységet.

Bár láttuk - legalábbis részben -, hogy az egyik ilyen eljárást hogyan lehet számítással elszámolni, még mindig számunkra mechanizmus van, amelyről nem tudunk elszámolni. Ezek közül néhányra a későbbi fejezetekben még visszatérünk. A 16. fejezetben további bizonyítékokat fogunk felvonultatni arra vonatkozóan, hogy a nyelvi viselkedés szabályvezérelt, és így számítással megvalósítható. A 19. fejezetben azt vizsgáljuk meg, hogyan modellezhetnénk néhány, az írott nyelv olvasásában szerepet játszó mechanizmust.

Egyelőre azonban fejezzük be ezt a fejezetet egy utolsó problémával. Úgy tűnik, hogy a nyelvi szövegértés egyik legnehezebben számon kérhető eleme a szemantika - a jelentésnek - a szintaktikai szerkezetekből való meghatározása. Vegyük például azt a tényt, hogy sok angol mondat kétértelmű.

Az amfibolizmus a mondatok olyan tulajdonsága, hogy szintaktikai felépítésükből adódóan egynél több szemantikai értelmezést tesznek lehetővé. Ez különbözik a lexikális kétértelműség eseteitől, amikor a homonímia többféle értelmezés lehetőségét vezet be. Az alábbi [8] mondat lexikailag kétértelmű, míg a [9] mondat kétértelmű:

[8] A bank kihívás elé állította a pilótát.

[9] Láttam az embert a dombon a távcsővel.

Míg a [8] mondat háromféleképpen értelmezhető a "bank" szó szemantikai értelmezésétől függően, addig a [9] mondat többféle értelmezést tesz lehetővé, de ez a kétértelműség nem függ a mondat bármelyik szavának kétértelműségétől. Inkább a mondat szintaktikai szerkezetét többféleképpen is értelmezhetjük, és ezek különböző jelentéseket eredményeznek.

A szemantika a szintaktikából való meghatározásának problémájára a 17. fejezetben térünk vissza részletesen.

EMBERI ÉRVELÉS

Az előző két fejezetben a racionális és a linguisztikus képességeket azzal a céllal közelítettük meg, hogy elemezzük az őket alkotó mechanizmusokat, és számítási szempontból számon tartjuk ezeket a mechanizmusokat.

Ebben és a következő fejezetben a pszichológia és a nyelvészet azon bizonyítékait fogjuk megvizsgálni, amelyek arra a kérdésre vonatkoznak, hogy az emberi racionalitás és a nyelvi kompetencia effektív módon szabályvezérelt és így számítással megvalósítható.

Az emberi racionalitással kapcsolatos empirikus adatok rövid vizsgálatában ebben a fejezetben segítségünkre lesz, ha először a következő gondolkodási feladatokra válaszolunk:

1. Ha Mike házas, akkor boldog. Mike házas. Ebből következik, hogy boldog?
2. Egyetlen hippi sem pénzügyi tanácsadó. Egyik pénzügyi tanácsadó sem atomtűntető. Ebből az következik, hogy néhány hippi atomtűntető?
3. Négy kártya van előtted. Mindegyik kártya egyik oldalán egy betű, a másikon egy szám van. A kártyák az egyik oldalukkal felfelé fekszenek, úgy, hogy a mutatott oldalakon a következőket láthatjátok:

AK47

- Azt mondják, hogy ez a szabály érvényes ezekre a kártyákra - Ha az *egyik oldalon egy A van, akkor a másik oldalon egy páros szám van*. A négy lap közül melyiket *kell felfordítanod*, hogy megállapítsd, hogy ez a szabály érvényes-e vagy sem?
4. Adriannek jó módú szülei vannak, és elit gimnáziumba járt. Befektetési bankárként dolgozik, egy vállalati vezető felesége, és Porschét vezet. Az alábbiak közül melyik a valószínűbb?

- (a) Adrian aggódik a közegészségügy és a közjólét miatt.
- (b) Adrian aggódik a közegészségügy és a jólét miatt, de konzervatíván szavaz.

5. Queenslandben minden ember kalapot visel a napon. Aki kalapot visel a napon, az nem kap melanomát az arcán. Ebből az következik, hogy Queenslandben senki sem kap melanomát az arcán?
6. Egy feldobott érme kilencszer egymás után fejre állt. Mennyi annak az esélye, hogy a következő dobásnál írás lesz?
- (a) Még a is.
(b) Jobb, mint a páros
(c) Kevesebb, mint egyenletes
7. Jonnak és Nicole-nak két gyermeke van. Az egyikük fiú. Mennyi az esélye, hogy a másik is fiú lesz?
- (a) 50% b)
33,33%
(c) 25%
8. Az egyetemen az összes vegetáriánus az Organic Food Cooperative tagja. Az egyetemen egyetlen vegetáriánus sem vásárol olyan élelmiszert, amelyről ismert, hogy növényvédőszerrel vagy tartósítószerekkel kezelték. Ebből következik, hogy a Bio Élelmiszer Szövetkezet tagjai közül senki sem vásárol olyan élelmiszert, amelyről ismert, hogy növényvédőszerrel vagy tartósítószerekkel kezelték?
9. Húzol egy kártyát egy normál pakliból. Az alábbiak közül melyik a valószínűbb?
- (a) Húzol egy ászt.
(b) Vörös ászt húzol.
10. Kapsz nyolc kártyát egy szabványos pakliból, és azt mondják, hogy a következő két állítás közül *csak* az egyik igaz:
- [1] Vagy egy ász vagy egy király van a kezedben (vagy mindkettő).
[2] Vagy egy ász vagy egy dáma van a kezedben (vagy mindkettő).

Az alábbiak közül melyik a valószínűbb?

- (a) Egy király van a kezedben.
(b) Egy ász van a kezedben.
11. Ha ebédidőben coffee-t iszik, akkor délután éberebb lesz. Dave ma délután nem éberebb. Ebből az következik, hogy Dave nem ivott coffee-t ebédidőben?
12. A munkanélküli-ellátások igénybevételére vonatkozó szabály kimondja: - *Ha ellátást vesz fel, akkor nem lehet*

a következő négy emberről kell további információkat gyűjtenie, hogy megbizonyosodjon arról, hogy nem szegik meg a szabályt?

- (a) Az a személy, akiről tudja, hogy alkalmazott.
- (b) Az a személy, akiről tudja, hogy munkanélküli.
- (c) Az a személy, akiről tudja, hogy ellátásban részesül.
- (d) Az a személy, akiről tudja, hogy nem részesül ellátásban.

15.1 LOGIKUSAN KÖVETVE

Elismert empirikus tény, hogy az emberek általában elég rosszul teljesítenek az érvelési problémák ezen csoportján. E kötet olvasója egyértelműen intelligensebb az átlagnál - már csak azért is, mert megvásárolta ezt a könyvet, ha másért nem -, de még így is meglepődnék, ha nem hibázna a feladatsoron (hacsak nem volt már korábban kitéve ezeknek a problémáknak).

Az a kérdés, hogy egy következtetés *logikusan következik-e* bizonyos premisszákból, a következtetés *érvényességére vonatkozóan* fogalmazódik meg. Egy következtetés akkor érvényes, ha a premisszáinak igazsága elegendő ahhoz, hogy garantálja a következtetés igazságát. Egy adott következtetés érvényessége annak a logikai formának az érvényességétől függ, amelyet az adott következtetés alkalmaz. Technikailag a következtetés logikai formái azok, amelyek érvényesek vagy nem érvényesek. Egy logikai forma érvényes iff nincs olyan példánya, amelynek igaz premisszái és hamis konklúziója van. Más szóval, egy logikai forma akkor érvényes, ha nincs olyan példánya, amely ellenpéldája az érvényességének.

Továbbra sem várom el, hogy jól értsd a logikai forma fogalmát, vagy hogy pontosan mit jelent egy logikai forma, de jelen célokból nem is szükséges, hogy értsd. Csupán arra akarok rámutatni, hogy az, hogy egy következtetés következik-e valamilyen premisszából - hogy a következtetés érvényes-e vagy sem -, tisztán formai megfontolás. A következtetések érvényességének meghatározásának egyáltalán semmi köze a premisszák és a konklúzió által kifejezett tényleges tartalomhoz vagy jelentéshez.

Bizonyítható azonban, hogy az *embereket* általában a premisszák és a következtetések jelentése vezérli, amikor a következtetések érvényességére vonatkozó tanulatlan döntéseket hozzák meg. Éppen ezért a kísérleti alanyok a fenti problémákban kiszámítható hibákat követnek el.

E fejezet elsődleges célja, hogy megvizsgáljuk, milyen hibákat követnek el az emberek általában az ilyen problémákban, és megvizsgáljuk, hogy a logikailag képzetlen alanyok racionális

158

teljesítménye a logikai problémákban kihívást jelent-e a komputacionalizmus számára.

15.2 RACIONÁLIS TELJESÍTMÉNY

Az emberek általában észreveszik az egyszerű következtetések érvényességét, mint például az 1. feladatban. Ez csak *modus ponens*, tehát ebből következik, hogy Mike boldog. A 11. feladat szintén egy nagyon egyszerű következtetési formát - *modus tollens* - tartalmaz. Egy igaz feltételes feltételes feltétel hamis következménye esetén mindig érvényesen következtethetünk az előfeltétel hamisságára. Azonban gyakori, hogy hibákat látunk ennél a feladatnál.

Ennek két oka lehet. Az egyik az, hogy a *modus tollens* tagadást tartalmaz, és úgy tűnik, hogy a tagadást is tartalmazó érvelés általában diffiktabb, mint a csak afirmálást tartalmazó érvelés. A másik ok az, hogy ahelyett, hogy a következtetés érvényességének meghatározásakor a premisszák igazságát feltételeznék, a reazonáló alanyok valószínűleg arra gondolnak, hogy Dave délutáni álmoságában valamilyen bonyolult tényező is szerepet játszhat - talán késő este volt -, és ezáltal megállapítják, hogy a következtetés nem következik. Más szóval, valószínűleg inkább a premisszák és a konklúzió jelentése, mint logikai formája fogja őket vezérelni.

Ez utóbbi megfontolás azonban valójában nem a következtetés érvényessége ellen szól, csupán az egyik premissza igazsága ellen. Az, hogy a premisszák ténylegesen igazak-e vagy sem, nem befolyásolja a következtetés érvényességét - egy következtetés akkor érvényes, ha a premisszák igazsága garantálja a következtetés igazságát. Az érvényes következtetéseknek lehetnek hamis premisszái, és a 11. feladat éppen ilyen eset - az ebédidőben történő koffeefogyasztás önmagában nem garantálja a délutáni éberséget, így a feltételes feltétel valójában hamis.

A 10. feladat további bizonyítékot ad arra, hogy a tagadás megnehezíti a reazoning feladatokat. A leggyakoribb válasz erre a feladatra a (b), de a válasz valójában az (a). A probléma információi alapján nem lehetséges, hogy van egy ász a kezünkben.

Azt mondják, hogy az [1] és a [2] közül *csak az egyik* igaz, ami azt jelenti, hogy az egyik állítás hamis. Ha az [1] hamis, akkor sem ász, sem király nincs a kezében. Ha a [2] hamis, akkor sem ász, sem dáma nincs a kezében. Tehát bármelyik állítás is bizonyul igaznak, nincs ász a kezében.

A tipikus hiba egyik magyarázata az, hogy az emberek általában figyelmen kívül hagyják a negatív információkat, és a pozitívakra koncentrálnak. Tehát ahelyett, hogy azon gondolkodnának, hogy az egyik állítás valótlansága mivel járna, inkább arra koncentrálnak, hogy bármelyik állítás igazsága mivel járna.

A tipikus hiba magyarázatának másik része az, hogy az, hogy

160

van egy ász a kézben, olyan dolog, ami mind a két esetben előfordul.

állítások, míg a király a kézben csak egy állításban szerepel. Ha tehát arra koncentrálunk, hogy az egyik állítás igazsága mit vonna maga után, akkor valószínűleg azt gondoljuk, hogy bármelyik állítás igazsága azt jelentené, hogy lehet ász a kézben, míg csak az egyik állítás igazsága azt jelentené, hogy lehet király a kézben, és így - tévesen - arra következtetünk, hogy valószínűbb, hogy van ász a kézben.

Az emberek általában elég rosszul teljesítenek a kategorikus gondolkodási feladatokban is, amint azt a 2., 5. és 8. probléma is mutatja. Ezek közül az 5. feladat a legegyszerűbb és a legvalószínűbb, hogy helyesen válaszolnak rá. A premisszákból ugyan következik, hogy Queenslandben senki sem kap arc melanomát, de még ebben az egyszerű esetben is tévútra vezethet bennünket, ha háttérismeretekkel rendelkezünk a bőrrák előfordulási gyakoriságáról azokon a helyeken, ahol erős napsütésnek vannak kitéve.

A 2. feladatra valószínűleg afirmatív válasz adható, de a helyes válasz az, hogy a következtetés ebben az esetben nem következik. A premisszák semmit sem árulnak el a hippik és az atomtüntetők közötti kapcsolatáról, így, bár nem következik, hogy az előbbiek egy része az utóbbi, de nem is kizárt. A tipikus hiba magyarázatának egy része itt a ~~negatív~~ kategorikus kapcsolatokról való gondolkodás diffikálására apellál, de a magyarázat fontos része az, hogy a sztereotip hippi nukleáris tüntető lenne, és ez a háttérinformáció akkor kerül előtérbe, ha a premisszák és a következtetés jelentését, nem pedig pusztán logikai formájukat vizsgáljuk.

A 8. feladatra is valószínű, hogy az afirmatív válasz adható, bár a helyes válasz az, hogy a következtetés nem következik. Ennek pontosan azok az okai, amelyek az 5. feladatnál is jellemző hibát okoznak. Ez részben azzal függ össze, hogy az érvelés negatív kategorikus kapcsolatokat tartalmaz, de nagyrészt azzal, hogy a Bioélelmiszer Szövetkezet sztereotip tagja kerülné az ilyen élelmiszereket.

A valószínűségről való gondolkodás szintén jellegzetes hibákhoz vezet, amint azt a 6. és 7. feladat mutatja. Azok, akik a 6. feladatra a "párosnál jobb" választ adják, a *szerecssejátékos tévedésének* áldozatává válnak. A szerecssejátékos tévedése az a nézet, hogy az esélyek "kiegyenlítődnek" a próbák bármelyik során. Ez természetesen hamis. Egy tisztességes pénzérme fejre esésének esélye mindig egyenlő, függetlenül a megelőző próbák számától. Tíz fej feldobása egymás után nem több és nem kevesebb a valószínűsége, mint a fej és írás bármely más kombinációjának feldobása.

A 7. problémára adott válasz - némileg ellentmondásos módon - az, hogy egy a háromhoz (33,33 százalék) az esélye annak, hogy a

162

másik gyermekük fiú lesz.

Nagyjából mindenki téved ebben a kérdésben. A válasz azért nem 50 százalékos, mert nem tudjuk, *hog*y melyikük gyermeke fiú, és ez affektálja a valószínűségi teret.

Négyféleképpen lehetséges, hogy Jonnak és Nicole-nak két gyermeke legyen. Lehet egy lányuk, majd egy fiuk, egy lányuk, majd egy másik lányuk, egy fiuk, majd egy lányuk, vagy egy fiuk, majd egy másik fiuk. Ha csak annyit tudunk, hogy az *egyik* gyermekük fiú - és nem tudjuk, melyik -, akkor az egyetlen kizárt helyzet az, hogy két lányuk lesz. Így három lehetőség marad, amelyek közül az egyik olyan, hogy a másik gyermekük is fiú, tehát ennek valószínűsége egy a háromhoz.

Ha tudnánk, hogy az *első* gyermekük fiú volt, vagy hogy a *második* gyermekük fiú volt, akkor ez az információ két lehetőséget kizárna, így a másik gyermek valószínűsége egy a kettőhöz lenne fiú. Így azonban csak annyi információnk van, hogy a négy lehetőség közül egyet kizárjunk.

A 4. és a 9. feladat különösen érdekes, mivel mindkét feladat szerkezetileg azonos, de a nem tanult megoldások tipikusan eltérnek egymástól. A 9. feladatra soha senki nem válaszol (b) választ, de a 4. feladatra gyakran válaszolnak (b) választ. Mindkét esetben a (b) a helytelen válasz, mivel egy konjunkció soha nem valószínűbb, mint bármelyik konjunkciója. Bár az emberek gyorsan felismerik, hogy az ás hűzése valószínűbb, mint a piros ás hűzése, a 4. feladatban leírt információkkal kapcsolatos háttértudásuk általában tévútra vezeti őket.

Úgy tűnik, hogy a 4. problémával kapcsolatos érvelési folyamat, amelyben az emberek részt vesznek, ismét a sztereotípiákra való hivatkozással jár. Egy tipikus befektetési bankár, aki elit gimnáziumi iskolát végzett és egy Porsche áll a garázsban, nem az a fajta ember, akitől elvárnánk, hogy aggódjon a közegészségügy és a közjólét miatt. Ha azonban ezen aggodalmak ellenére hajlamos konzervatívan szavazni, akkor ez közelebbi - bár még mindig kissé rendhagyó - illeszkedést jelent a sztereotípiához. Így hajlamosak vagyunk azt gondolni, hogy a (b) a valószínűbb eset, tekintve, amit Adriánról mondtak nekünk.

A 3. és a 12. feladat azért is különösen érdekes, mert szerkezetileg ezek is azonosak - pontosan ugyanaz a logikai forma -, mégis jellemzően különböző válaszokat kapnak. A 3. feladatra a leggyakoribb válasz az, hogy fel kell fordítanunk az "A"-t és a "4"-et mutató kártyát. A helyes válasz azonban az, hogy az "A"-t és a "7"-et mutató kártyát kell felfordítanunk.

Ennek az az oka, hogy ahhoz, hogy megállapítsuk, hogy egy szabály érvényes-e vagy sem, nem megerősítő, hanem cáfoló eseteket kell keresnünk.

esetek. Ellenpélda hiányában azt mondhatjuk, hogy a szabály érvényes - ha találunk ellenpéldát, akkor bizonyíthatjuk, hogy nem érvényes.

A problémás esetben nem kell felfordítanunk a '4'-es kártyát. Ha a másik oldalon ász van, akkor ez csak megerősíti a szabályt, de ha nincs, akkor nem ad ellenpéldát. A szabály csak annyit mond, hogy ha az egyik oldalon egy ász van, akkor a másik oldalon egy páros szám van. Nem mondja ki, hogy minek kell lennie, ha az egyik oldalon páros szám van.

Az "A"-t mutató kártyát azonban meg kell fordítanunk, hogy megbizonyosodjunk arról, hogy a másik oldalon páros szám van. A "7"-est mutató kártyát is fel kell fordítanunk, hogy megbizonyosodjunk arról, hogy a másik oldalon *nem* egy "A" van, mivel ez a szabály ellenpéldája lenne, és így megmutatná, hogy a szabály nem érvényes.

Míg a 3. problémát szinte mindig elrontják, addig a 12. problémát szinte mindig jól csinálják, pedig ez pontosan ugyanaz a probléma. A 12. feladatra adott helyes válasz az, hogy ellenőriznünk kell a foglalkoztatott személyt - hogy megbizonyosodjunk arról, hogy nem vesz-e igénybe ellátást - és a segélyt igénybe vevő személyt - hogy megbizonyosodjunk arról, hogy nem foglalkoztatott-e szintén.

Ezt a performatív ellentétet a strukturálisan identikus érvelési feladatokban úgy tudjuk megmagyarázni, hogy ismét a következtető alany háttérismeretére hivatkozunk a probléma információi tekintetében. Legtöbbünk pontosan tudja, hogy mi a jóléti csaló, és implicit módon megérti, hogy az olyan valaki, aki megszegi a szabályokat. Következésképpen pontosan tudjuk, hogy mit kell keresnünk a 12. kérdésben - a szabályszegés lehetséges eseteit.

Az egyik oldalon számokkal, a másik oldalon betűkkel ellátott kártyák ezzel szemben nem olyasmi, amivel a legtöbben közülünk valaha is találkozónának. Mint ilyen, nincs olyan releváns háttérinformáció, amely azt mondaná, hogy a 3. feladat megoldásához a szabályszegés eseteit kellene keresnünk.

15.3 LELKI MODELLEK

Úgy tűnik, rengeteg bizonyítékunk van arra, hogy amikor az emberek érvelnek, általában nem kifejezetten formális szabályokat követnek. Inkább mentális modelleket alkotnak a problémás helyzetről, és ezeket a mentális modelleket kérdezik ki a megoldás meghatározása érdekében.

Ezek a mentális modellek néha teljes mértékben - és szelektíven - a problémában megadott információk alapján épülnek fel, de gyakran hivatkoznak releváns háttérinformációkra és sztereotip vagy *paradigmatikus* esetekkel való összehasonlításokra is.

A 2. feladat megválaszolásakor például a mentális modell, amit felépítünk, egy hippit tartalmaz, aki nem befektetési bankár, és hajlamosak vagyunk

azt a kérdést, hogy *néhány* ilyen hippi atomtűntető-e, úgy értelmezzük, mint azt a kérdést, hogy valószínű vagy lehetséges-e, hogy a paradigma ilyen hippi atomtűntető legyen. Következésképpen tévesen afirmatív választ adunk.

Hasonlóképpen, a 4. kérdés megválaszolásához egy kiváltságos háttérrel és kiváltságos életmóddal rendelkező befektetési bankár mentális modelljét állítjuk fel. Ezután mérlegeljük az ilyen személy paradigmáját társadalmi aggodalmak valószínűségével szemben, és azt találjuk, hogy míg az ilyen társadalmi koncepció nem illeszkedik a sztereotípiához, addig a konzervatív szavazási preferencia hozzáadása némileg közelebbi illeszkedést eredményez. Következésképpen tévesen azt gondoljuk, hogy az a modell a valószínűbb, amelyik konzervatív szavazási preferenciát is tartalmaz.

Bár ezeknek a mentális modelleknek a felépítése néha tévútra vezethet minket, néha nagyon gyorsan elvezethetnek a helyes válaszhoz, mint a 12. feladatban. Evolúciós szempontból elvárható, hogy az emberek olyan érvelési eljárásokat fejlesszenek ki, amelyek a lehető legkevesebb kognitív erőforrást igénylik, és amelyek fontosnak tartják a hasonló helyzetekben szerzett korábbi tapasztalatokat. Így várható, hogy az emberek gyakran figyelmen kívül hagynak bizonyos problémainformációkat annak érdekében, hogy ahol csak lehet, egyszerűsítsenek, és hogy látszólag releváns háttérinformációkra hivatkoznak, még akkor is, ha szigorúan csak a probléma formális tulajdonságai relevánsak.

Ha ez azt jelenti, hogy néha rosszul teljesítünk a meglehetősen mesterséges formális feladatokban, akkor ez egy kis evolúciós ár azért, hogy a valós világbeli gondolkodási feladatokban gyorsan és olcsón (kognitív értelemben) a legtöbbször jól végezzük a dolgunkat.

Ez az állítás, miszerint a tipikus emberi gondolkodás mentális modellek felépítésével és kikérdezésével jár, az egyik olyan elmélet, amely összefügg és magyarázatot ad az ilyen gondolkodási feladatokban nyújtott tipikus teljesítményre. Bár ez csak egy elmélet, mégis eléggé intuitív, és a kognitív pszichológiában így vagy úgy és különböző elnevezések alatt némi népszerűségnek örvend. Az ilyen típusú elméletnek, ahogyan az ezekre az érvelési feladatokra vonatkozik, Johnson-Laird, Tversky és Kahneman voltak az elődei.

15.4 MAGYARÁZÓ TEHER

Ha az lenne a helyzet, hogy az emberek mindig formálisan érvelnek, valamilyen logika diktátumai szerint, akkor a racionális mechanizmusok számításbeli bemutatása nagyon egyszerű lenne,

mivel a logikák egyszerűen formális rendszerek.

Az érvelési feladatokra vonatkozó tipikus teljesítményre vonatkozó empirikus adatok fényében azonban a számítástudományi szakembernek szembe kell néznie azzal a kérdéssel, hogy

magyarázó kihívás - a mentális modellek felépítésének és interrogációjának számítási szempontok szerinti magyarázata.

Nincs azonban semmilyen prima facie ok, és az empirikus adatokból sem tudunk olyan okot levezetni, amely arra engedne következtetni, hogy ezek az érvelési mechanizmusok nem számíthatók ki számítással. Ha nem lennének filozófiai szempontból óvatosak, kísértésbe eshetnének, hogy engedélyezzünk egy olyan érvet a komputacionalizmus ellen, mint a következő.

A tanulatlan alanyok érvelési folyamatai bizonyíthatóan nem egyszerűen a logikai szabályok explicit követését foglalják magukban.

Az emberek jellemzően logikátlanul, vagy irracionálisan érvelnek.

Ezek az érvelési mechanizmusok nem számolhatók el a formális rendszerek szempontjából.

Van legalább egy olyan mentális folyamat, amely nem valósítható meg számítással.

A komputacionalizmus hamis

A fenti érvelés már az első következtetésnél rosszul kezdődik. Az "irracionálisnak" vannak erős és gyenge értelmezései, amelyekkel kapcsolatban óvatosnak kell lennünk, hogy ne tegyünk egyenlőségjelet. A gyenge értelemben vett "irracionális" akkor vádolunk valakit, ha érvelését olyan elv vezérli, amely valójában hamis. Aki a szerencsejátékos tévedésének áldozatául esik, az a gyenge értelemben vett irracionális gondolkodás paradigmatis esete.

Az "irracionálisnak" ez a gyenge értelme azonban nem elég erős ahhoz, hogy a komputacionalizmus elleni érvelés következő következtetését igazolja. Az "irracionálisnak" az az értelme, amelyet a tanulatlan alanyoknak tulajdonítunk, és amely alapján a fenti érvelés következő lépését elfogadnánk, sokkal erősebb értelem. Ebben az erősebb értelemben akkor vádolnánk valakit "irracionálisnak", ha érvelését egyáltalán nem vezérli semmilyen elv, vagy ha egy olyan érvelés ellenére, amely során minden premissza igazságát és az érvelés érvényességét elfogadja, mégis elutasítja a következtetés elfogadását.

Nem az a helyzet azonban, hogy a tipikus alanyok érvelési teljesítménye olyan, hogy az "irracionális" eme erős értelemben vett vádját indokolná. Minden esetben jó bizonyítékok szólnak amellet, hogy a bevetett érvelési folyamatokat valójában bizonyos

elvek - csak éppen nem olyan formális elvek, amelyek garantáltan igazságtartóak.

Továbbá, ha elmagyarázzák a helyes érvelési módszereket azoknak az alanyoknak, akik hibásan döntöttek az érvelési feladatokban, akkor általában elég gyorsan észreveszik a hibájukat, és nem valószínű, hogy a jövőbeni feladatokban újra elkövetik ugyanazt a hibát. Kezdetben ellenállhatnak a helyes következtetés elfogadásának - különösen a 3. és a 7. feladatnál -, azonban amint megfelelően elmagyarázzák az érvelést (esetleg ábrákkal vagy analóg helyzetekre való hivatkozással), ez az ellenállás leküzdhető.

Következésképpen az empirikus adatok nem elegendőek ahhoz, hogy igazolják azt az erős állítást, hogy az emberi gondolkodási mechanizmusok *nem* számolhatók el formális rendszerek segítségével - egy olyan állítás, amelynek igazsága a komputacionalizmus hamisságát bizonyítaná.

Az emberi gondolkodási mechanizmusokat természetesen nem lehet pusztán a formális szabályok *explicit követésével* magyarázni, de ez nem jelenti *ipso facto* azt, hogy a mechanizmusokat ne irányítanák implicit módon számítási módszerek. Hasonlóképpen, a mondatok grammatikájának megítélése nem foglalja magában a formális szabályok *explicit* követését, de ez nem bizonyítja, hogy az ilyen megítéléseket nem számítási folyamatok írják alá.

Az empirikus adatok által a komputacionalizmus számára támasztott kihívás tehát nem leküzdhetetlen, és arra a magyarázó teherre korlátozódik, hogy a tipikus reakciókban részt vevő mechanizmusok számításal történő bemutatását kell ~~de~~

Úgy tűnik, hogy az egyik központi mechanizmus, amely az ilyen érvelésben szerepet játszik, az a képesség, hogy összehasonlításokat tudunk végezni a múltbeli helyzetekkel és a hasonló esetekkel. Ez magában foglalja annak felismerését, hogy a problémás helyzet egy ismert tapasztalati mintát tartalmaz, és ennek az ismert mintának a felhasználását a feladat szempontjából releváns információk meghatározásához.

Ez a mintaillesztés és rekonstrukció egy olyan mechanizmus, amelyet talán nagyon jól tudunk elszámolni azokkal a rendszerekkel, amelyeket a 19. fejezetben fogunk megvizsgálni.

EMBERI NYELV

Az emberi nyelvi kapacitás valóban egészen elképesztő. A nyelvi produkciót és megértést elősegítő mechanizmusok meglepően összetettek, tekintve, hogy a nyelvi képességünk annyira természetes számunkra, hogy hihetetlenül egyszerűnek tűnik. Kétségtelen, hogy a 14. fejezet elolvasása után elkezdte értékelni, hogy a nyelvi viselkedés mennyi kognitív feldolgozással jár.

Bármennyire is diffikult és összetett a nyelvi képességek elmélete, egy hatéves gyermek már teljesen internalizálta első nyelvének szintaxisát, morfológiáját és fonológiáját. Ez arra készítette a nyelvészkatatókat, hogy Chomsky nyomán valamilyen veleszületett mechanizmus szükségességét tételezzék fel, amely segíti első nyelvünk elsajátítását.

Sok mindent el lehet mondani erről a feltételezett veleszületett mechanizmusról és annak a nyelvésajátításban betöltött szerepéről, de ez a vita nagyon kevésbé érinti a komputualizmus tarthatóságát, ezért ebben a fejezetben nem erre szeretnék összpontosítani.

Ennek a fejezetnek az a célja, hogy bizonyítékot találjon arra, hogy nyelvi tevékenységünk nagy része szigorúan szabályozott - és így számítással megvalósítható - annak ellenére, hogy nem ismerjük ezeket a szabályokat. Ez remélhetőleg súlyt ad az előző fejezet azon állításának, hogy még azok a kognitív mechanizmusok is számítással számolhatók, amelyek nem tartalmaznak explicit szabálykövetést.

Számos bizonyíték szól amellet a nézet mellett, hogy a nyelvi viselkedés teljes mértékben szabályvezérelt. A nyelvészeti kutatások nagy része ezeknek az implicit szabályoknak a kifejezésére és kodifikálására irányul.

A nyelvészeti tanulmányok egyik olyan területe, ahol a viselkedésünket irányító implicit szabályok azonosítása jól kimutatható, a *fonológia* tanulmányozása.

16.1 OBSTRUENS FONÉMÁK

A fonológia a beszélt nyelv beszédhangjainak és hangmintáinak tanulmányozása. A fonológia tanulmányozásának középpontjában egy adott nyelv fonémáinak azonosítása és osztályozása áll. A fonémákról a 14. fejezetben tanultunk egy kicsit, ahol láttuk, hogy a fonémák a beszéd legkisebb egységei, amelyek megkülönböztető kontrasztot biztosítanak. Most erre az ismeretre építünk, és tax- az angol nyelv fonémáinak onomizálása.

A fonémák nyílt vagy *szonoráns* hangokra - amelyeket magánhangzónak tekinthetünk - és korlátozott vagy *obstruens* hangokra - amelyeket mássalhangzónak tekinthetünk - oszlanak.

Az obstruens fonémákat az artikuláció helyének és módjának, valamint annak alapján írjuk le, hogy hangzottak-e vagy sem. Az artikuláció helye az artikulációs apparátusnak arra a kombinációjára utal, amelyet a produkciójuk során alkalmaznak. Az artikuláció módja arra utal, hogy a hangot milyen mértékben korlátozza az artikulációs apparátus. A 16.1. ábrán látható táblázat az obstruens fonémákat a függőleges tengely mentén a hangképzés és az artikuláció módja szerint osztályozza, és a vízszintes tengely mentén lévő csuklójuk helye.

Mielőtt tovább olvasna, nézze meg a 16.2. ábrán található kiejtési táblázatot, és gyakorolja az egyes fonémák kiejtését, figyelve az ajkak, a fogak és a nyelv elhelyezkedésére, valamint arra, hogy az egyes fonémák kiejtésénél ez az elhelyezkedés milyen mértékben korlátozza a levegő átjutását.

	Bilabialis	Labiodentális	Interdentális	Alveoláris	Alveopalatális	Veláris	Glottális
Hangos megállások	b			d	g		
Hang nélküli megállók	p			t	k		
Orrlyukak	m			n			ŋ
Hangos frikatívumok	v	ð	z	s			
Hangtalan frikatívumok	f	θ	ʃ	h			
Affrikáták				tʃ			
Hang nélküli affrikáták				ts			
Hozzávetőlegések	w			r			l

16.1. ábra Az angol obstruens fonémák taxonómiája.

b	bat	ʒ	asian
d	dot	f	fit
g	got	θ	thin
p	pot	s	sat
t	tot	ʃ	shy
k	cat	h	hot
m	mat	dʒ	jury
n	net	tʃ	chat
ŋ	sing	w	wet
v	vet	r	rat
ð	that	l	lot
z	size	j	yacht

16.2. ábra Az angol obstruens fonémák kiejtési táblázata.

A stopok vagy plosívák olyan fonémák, amelyek előállítása a levegőnek az artikulációs készüléken való áthaladásának teljes elzárásával jár. A ploszívumok lehetnek hangosak, mint például /d/, vagy hangtalanok, mint például /t/.

A frikatívok olyan fonémák, amelyek előállítása során a levegőnek az artikulációs készüléken való áthaladása korlátozódik, így a levegő áthaladása sziszegő, sziszegő hangot eredményez. A frikatívumok lehetnek hangosak, mint például az /z/, vagy hangtalanok, mint például az /s/.

Az orrhangok azok a hangok, amelyek előállítása a levegőnek az orrüregben való rezonálásával jár. Ezek azok az obstruáló hangok, amelyeknek a keletkezése megfázáskor gátolt. Ha befogjuk az orrunkat, miközben megpróbáljuk ezeket a hangokat produkálni, azt fogjuk tapasztalni, hogy a hang lapos és fapados lesz, míg az orr befogása nincs hatással más obstruens hangok produkációjára. Minden orrhangzó fonéma hangzó.

Az affricate fonémák plosívumok és frikatívumok kombinációi. A levegő áthaladása kezdetben teljesen, majd részben korlátozott. Az angolban csak két affricate van - a hang nélküli affricate, amely a "chat" szó kezdőhangja, és a hang nélküli affricate, amely a "fridge" szó záróhangja.

Végül, vannak approximánsok. Ezeknek a fonémáknak az osztályozása némileg vitatott, de a fenti osztályozás a mi céljainkat szolgálja. Minden közelítő hangzó hangzó. A közelítő fonémák a szonoráns fonémákhoz legközelebb álló obstruensek.

Ez kimeríti az obstruens fonémák artikulációs módját.

Az osztályozás másik tengelye az artikulációs hely, amely a productionben részt vevő elsődleges artikulációs apparátust írja le.

A bilabiális fonémák azok, amelyek előállításához mindkét ajkát használni kell. A labiodentális fonémák a felső fogak alsó ajkához való illesztését foglalják magukban. Az interdentalis fonémák a nyelvnek a fogak közé helyezését jelölik.

Az alveoláris gerinc a felső fogak mögött elhelyezkedő kemény, kiemelkedő gerinc. Az alveoláris fonémák azok, amelyek előállítása során a nyelv az alveoláris gerinchez vagy annak közelébe kerül.

Az alveoláris gerinc mögött a felső állkapocs mentén folytatódik a kemény szájpadlás. Az alveopalatis fonémák úgy jönnek létre, hogy a nyelvet az alveoláris gerinc mögé és a kemény szájpadláshoz vagy annak közelébe helyezzük.

Ha végighúzza az ujját a szájpadláson a torok hátsó része felé, a kemény szájpadlás végéig, akkor érezni fogja a lágy szájpadlás puha részét. A lágy szájpadlást velumnak is nevezik. A veláris fonémák azok, amelyeket a nyelvnek a velum közelébe vagy a velumhoz való helyezését jelölik.

A glottis a hangszalagok közötti tér. A glottális frika- tív /h/ a levegőnek a glottison való áthaladásának korlátozásával, de a hangszalagok rezgése nélkül jön létre.

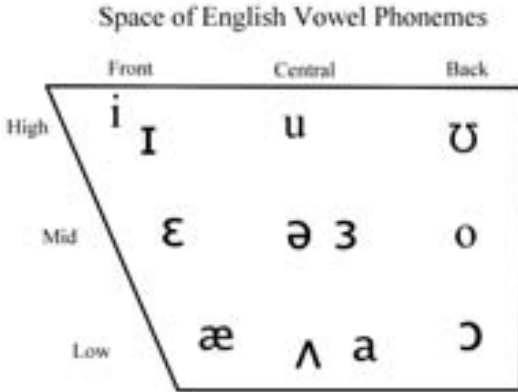
Az angol nyelv egyes nyelvjárásainak beszélői - például a londoni cockney dialektus - bizonyos fonémás kontextusokban (például a "bottle" vagy a "sorted" szavakban) az alveoláris stop helyett glottal stopot használnak. Ez nem egy különálló angol fonéma, hanem a /t/ fonéma *allofón változata*. Az allofonokról hamarosan többet fogunk megtudni, de először is taxonómizáljuk a szonoráns (magánhangzó) fonémákat.

16.2 SZONORÁNS FONÉMÁK

Minden szonoráns fonéma *ipso facto* hangos. Az artikulációjuk is azonos - nyílt hangok. Más szóval a levegőnek az artikulációs készüléken való áthaladása nem akadályozott, hanem szabadon rezonál a szájüregben.

A szonoráns fonémákat két tengely mentén taxonómizáljuk, amelyek a nyelvnek a szájban a **prókó** során elfoglalt helyzetét írják le. Az egyik tengely a nyelv magasságának felel meg, a másik pedig a nyelv felemelt vagy leeresztett részének.

Ellentétben az obstruens fonémákkal, amelyek a fonémák között egyértelmű különbségeket tesznek lehetővé, mivel a termelésükben részt vevő artikulációs apparátus különálló, a szonoráns fonémák a fonémák között egy különálló artikulációs apparátusban vannak.



16.3 ábra Magánhangzótér.

kontinuum, mivel a nyelv lehetséges helyzete a szájban folyamatos, nem diszkrét. Mint ilyenek, a magánhangzó-fonémák a lehetséges magánhangzók kétdimenziós terébe esnek.

Következésképpen a 16.3. ábrán látható módon *kardinális* magánhangzókat azonosítunk ezen a magánhangzó-téren belül. A magánhangzó-tér a legmagasabb, legfrontálisabb szonoráns fonémától a legalacsonyabb és legkevésbé frontális szonoráns fonémáig folyamatos.

A 16.3. ábrán bemutatott kardinális magánhangzók mind *egyhangúak*. Ez azt jelenti, hogy előállításuk egyetlen folyamatos nyelvállással történik. A monoftongus fonémák mellett azonban vannak *diftongus* fonémák is. A diftongok olyan fonémák, amelyek produkciója az egyik kardinális magánhangzó pozíciójában kezdődik, de a hang produkciója során egy másik kardinális magánhangzó nyelvhelyzete felé mozog.

A 16.4. ábrán látható kiejtési táblázat az általam beszélt ausztrál angol nyelvjárás magánhangzóit mutatja be. Megjegyzendő azonban, hogy a magánhangzók kiejtése az angol nyelvjárások között változik, különösen a diftongusok, de egyes monofonhangok kiejtése tekintetében is. Következésképpen a 16.4. ábrán látható hármashangzók közül néhány, amelyek az én angol nyelvjárásomban mind rímelnek, az Ön nyelvjárásában nem biztos, hogy rímelnek, attól függően, hogy Ön hol tanult angolul beszélni. Emellett az ausztrálon kívüli angol nyelvjárások - különösen az amerikai nyelvjárások - beszélői valószínűleg monoftongusokat használnak néhány diftongus helyett.

Monophthongs

i	pea	feet	ease
I	pit	fit	is
ɛ	pet	fret	met
æ	pat	fat	mat
u	put	foot	should
ʊ	lute	fruit	shoot
ə	pencil	the	effort
ɜ	purred	heard	shirt
o	pot	hot	shot
ʌ	putt	butter	mother
a	part	barter	father
ɔ	port	bought	fraught

Diphthongs

aI	l	my	sigh
ɛI	ate	weight	shave
ɔI	toy	oil	buoy
Iə	here	hear	seer
ɛə	there	share	bear
ʊə	lure	tour	fewer
aʊ	loud	now	pout
oʊ	boat	stow	hoe

16.4. ábra Az ausztrál angol magánhangzófonémák kiejtési táblázata.

16.3 ALLOFONOK ÉS FONETIKAI MEGVALÓSÍTÁS

A fonémák idealizálások. A tényleges beszédhangok - a *telefonok* - megközelítőleg megfelelnek a fonémáknak, és beszélőnként jelentősen eltérhetnek egymástól. Nyelvhasználóként nagyon jók vagyunk a beszédhangok megkülönböztető kontrasztjainak felismerésében és a telefonok fonémákhoz való hozzárendelésében.

Az első érdekes pont itt az, hogy a tipikus egynyelvű beszélők csak az anyanyelvükben található fonémákhoz tudnak telefonokat asszimilálni. Ez azt jelenti, hogy szó szerint nem hallanak olyan fonémákat, amelyek nem szerepelnek a nyelvükben, mivel nem érzékenyek az anyanyelvükben nem jellemző fonetikai kontrasztokra.

Mivel a természetes nyelvek különböznek a szemantikailag megkülönböztető fonetikai kontrasztokban, ez különösen diffiktté teheti egy második nyelv tanulását, ha a második nyelv olyan fonémákat tartalmaz, amelyek az első nyelvben nem szerepelnek. A vietnámi és a mandarin nyelvben például a beszédhangok hangszíne megkülönböztető. A hangszín azonban az angolban nem disztinktív kontraszt, így az ezeket a nyelveket tanuló angol anyanyelvűek számára óriási diffikrumot jelent a disztinktúra hallása, mivel a megfelelő hangokat automatikusan ugyanahhoz az angol fonémához asszimilálják.

Ezt bonyolítja az a tény, hogy egy és ugyanaz a fonéma különböző (de szemantikailag nem megkülönböztethető) *allofonokban* is megvalósulhat a hangtani kontextustól függően. Az allofonok közötti fonetikai differencia a nyelv anyanyelvi beszélője számára nagyon difficult, mivel a kontraszt nem megkülönböztető, hanem az előfordulás kontextusából kiszámítható.

Az allofonok egymást kiegészítő eloszlásban fordulnak elő. Ez azt jelenti, hogy ha egy fonémához egynél több allofon társul, akkor minden allofon mindig és kizárólag a *fonetikai megvalósulási szabályoknak* megfelelően, kiszámítható fonémakon- textusokban fog előállni. Az allofonok hangtani megvalósulása komplementer eloszlásban *szigorúan szabályozott*. Néhány példa ezt szemléletesen világossá teszi.

Az angol nyelv egyik fonetikai megvalósítási szabálya a hang nélküli megálló *aspirációját* foglalja magában, amikor azok szókezdetűek. Ez azt jelenti, hogy ha egy szó hang nélküli végződéssel kezdődik, akkor a végződés egy kis puff levegővel keletkezik, ami más kontextusban *nem* keletkezik.

Tegye a tenyerét közvetlenül a szája elé, és mondja ki többször a "top" és a "stop" szavakat, és észre kell vennie a különbséget. Amikor a 'top' szó /t/ kezdetű szavát mondja, éreznie kell a tenyerén a levegő extra puffját. Amikor a

/t/ a "stop"-ban, azonban ez a levegő puffja hiányzik, mivel az /t/ allofonja nem aspirált. Próbáljuk ki ezt más olyan szópárokkal, amelyek tartalmaznak

aspirált és nem aspirált hang nélküli stop allofonok, mint például a /k/ a 'Cate'-ben és a 'skate'-ben vagy az /p/ a 'pat'-ben és az 'alma'-ban.

Nagyon difficult azonban ténylegesen hallani a differenciát e hangok között, pontosan azért, mert allofón változatokról van szó - az aspiráció nem megkülönböztető kontraszt az angolban.

Az angol nyelv másik hangtani megvalósítási szabálya, hogy a magánhangzók mindig és kizárólag a nazális obstruensek előtt nazalizálódnak. Tartsa az orrát, miközben többször kiejti a 'cat' és a 'can' szót, és meg kell tudnia állapítani a különbséget. Próbálja ki ezt más, nazalizált és nem nazalizált magánhangzókat tartalmazó szópárokkal is, mint például a "pot" és a "pond" vagy a "sit" és a "sing".

Egyes fonetikai megvalósítási szabályok csak bizonyos nyelvjárásokra korlátozódnak, mint például az a szabály, amely a /t/ glottal stop allofonjának előállítását szabályozza a cockney-i angolban. A cockney angolt beszélők még mindig a standard hang nélküli alveoláris stop allofónját produkálják, amikor az /t/ szókezdő vagy frikatívumot követ, mint például a 'tell', 'start', 'faster' vagy 'softer' szavakban, de a glottális stopot produkálják, amikor az /t/ intervokális (magánhangzók közötti) környezetben fordul elő, mint például a 'butter' vagy 'letter' szavakban, és akkor is, amikor az /t/ szó végén fordul elő, mint például a 'short' vagy 'let' szavakban.

A mi céljaink szempontjából az az érdekes, hogy ez a szigorúan szabályozott tevékenység annak ellenére történik, hogy nem ismerjük ezeket a fonetikai megvalósítási szabályokat, és teljesen érzéketlenek vagyunk az előállított allofonok közötti fonetikai különbségek iránt.

Valójában ezek a fonetikai megvalósítási szabályok olyan mélyen gyökereznek, hogy nem tehetünk róla, hogy egy második nyelv elsajátításakor illegálisan átvesszük őket. Általában képtelenek vagyunk észrevenni, hogy valójában differens hangokat produkálunk differens kontextusokban, de ez valószínűleg nyilvánvaló lesz a tanult nyelv anyanyelvi beszélője számára.

Erre példa, hogy a franciát tanuló angol anyanyelvűek továbbra is (helytelenül) aspirálják a szavak hang nélküli kezdőhangjait - például a /t/-t a "Tour de France"-ban - még akkor is, ha erre felhívták a figyelmüket. Egy olyan nyelv beszélőiként, ahol ez az aspirált fonéma más allofonikus változatokkal kiegészítő eloszlásban van, egyszerűen képtelenek vagyunk észrevenni, hogy ezt tesszük. Hasonlóképpen az angolul tanuló német anyanyelvűek hajlamosak arra, hogy a szóvégződéseket devoiceálják - például a /g/-t a 'dog'-ban /k/-ként adják ki.

Tehát nemcsak arra törekszünk, hogy a második nyelvünk hangjait kizárólag az első nyelvünk fonémáinak felhasználásával próbáljuk meg előadni, hanem az anyanyelvi fonetikai megvalósítási

szabályainkat is átvisszük a második nyelvünkbe. Ez a két tényező felelős az akcentusért, amitől egy nem anyanyelvi beszélő jellemzően soha nem fog tudni teljesen megszabadulni. Sőt - az adott első és második nyelvtől függően - még az is előfordulhat, hogy

képesek a második nyelv fonémakészletének pontos előállítására, de a fonetikai megvalósítási szabályok felülírása nagyon diffikult. Ez a helyzet az azonos nyelv dialektikus változatai esetében is, mint például az angol nyelv számos dialektusa.

Szerencsére az akcentusok feldolgozásában is elég jók vagyunk általában. Néha kell némi kitettség, mire képesek leszünk az akcentusokat könnyedén feldolgozni - én a glasgow-i akcentust először áthatolhatatlannak találtam -, de ha sokáig kiteszük magunkat, implicit módon megtanulunk egy szűrőt alkalmazni a hallott hangokra, amely a szándékolt fonémákhoz asszimilálja őket. Bár kezdetben lehet, hogy kifejezetten erre kell összpontosítanunk, gyorsan belsővé tesszük ezeket az átalakító elveket.

Ez azt jelzi, hogy a szabályok által vezérelt nyelvi elveket még jóval az anyanyelv elsajátítása után is internalizáljuk, még akkor is, ha soha nem tanulunk második nyelvet. Amikor először találkozunk az angol nyelvnek a miénktől eltérő dialektusú beszélőivel vagy az angol mint második nyelv beszélőivel, nagyon gyorsan internalizáljuk azokat a szabályokat, amelyek lehetővé teszik számunkra, hogy könnyedén feldolgozzuk az akcentusukat. Természetesen, ami az azonos nyelv egy másik dialektusának beszélőjét illeti, az akcentus a miénk - biztos vagyok benne, hogy a glasgow-iak az én ausztrál akcentusomat ugyanolyan áthatolhatatlannak találják kezdetben, mint én az övékét.

16.4 ELSŐ NYELVTANULÁS

Az első nyelvelsajátítási szakirodalomban is számos bizonyítékot találunk a szabályvezérelt nyelvi aktivitásra.

Amikor a gyermek eléri a tizenkét hónapos kort, a nyelvtanulás egyszavas szakaszába lép. Képesek arra, hogy egyszavas kifejezésekkel mutassanak tárgyakra és megnevezzék azokat, és néhány vágyukat is képesek egyszavas kifejezésekkel jelezni - "labda", "baba", "mami".

Ebben a szakaszban gyakran előfordul, hogy az újonnan szerzett kifejezések szemantikai *túltágulása* és szemantikai *alultágulása* egyaránt megfigyelhető. Túlterjeszkedésről akkor beszélünk, amikor egy kifejezést a megfelelő kiterjesztésén túlmutató referensekre alkalmazunk. Amikor például egy gyermek először elsajátítja a "labda" szót, ezt a kifejezést más kerek tárgyakra is alkalmazhatja, például gyümölcsökre vagy a Holdra. Hasonlóképpen, ha van egy kutyája, és megtanulja a nevét, akkor ezt a kifejezést minden kutyára alkalmazhatja.

A szemantikai alulbővítés, mint azt bizonyára már kitalálta, ennek az ellenkezője. Az alulterjeszkedés akkor következik be,

amikor a gyermek egy újonnan elsajátított kifejezés alkalmazását csak bizonyos megfelelő fogalmakra korlátozza.

referensek. Előfordulhat például, hogy a "játék" szót csak egy bizonyos kedvenc játékkra használják. Az alulfeszítés általában ritkábban fordul elő, mint a túlfeszítés.

Mindkét esetben a gyermek viszonylag gyorsan megtanulja az új kifejezés helyes alkalmazási körét a lexikonjában. Ez legalábbis *prima facie* bizonyíték arra, hogy a gyermek internalizálja a kifejezés helyes alkalmazását szabályozó szabályokat, amelyek a referensek megfigyelhető jellemzőin alapulnak.

Az egyszavas szakasz másik tipikus jellemzője bizonyos fonémák szisztematikus fonémahelyettesítése, amelyeket a gyermek nem tud előállítani. Például az alveopális hangtalan frikatívum eléggé diffikultuszi fonéma, mivel meglehetősen ügyes nyelvhelyzetet igényel, ezért nem ritka, hogy a gyerekek szisztematikusán vagy a hangtalan alveoláris frikatívummal, vagy a hangtalan interdentalis frikatívummal helyettesítik - így a "ship" kiejtése "sip" vagy "thip".

E fonémahelyettesítések mellett a gyermek valószínűleg szisztematikus fonológiai egyszerűsítéseket is végez. Az "alvás" kezdőbetűje fonetikai szempontból is nehéz, így a gyermek valószínűleg a "szivárog" szót fogja előállítani helyette. A fonológiaiailag összetettebb szavakat még tovább egyszerűsítik - gyakran sajátos módon. Gyakori példa erre a "sketty" a "spagetti" szóra.

A különösen érdekes pont itt nem csak az, hogy a gyerekek szisztematikusán végzik ezeket a helyettesítéseket, hanem az is, hogy érzékenyek arra a tényre, hogy ezt teszik. Jellemzően tökéletesen képesek meghallani a különbséget a "seep" és az "sleep" felnőtt általi kiejtése között, és életkoruktól függően rájöhetnek, hogy csúfolják őket, ha a helyes kiejtés helyett az ő fonetikus kimenetüket adják vissza. A fonémák megértésének képessége tehát gyorsabban fejlődik, mint a fonémák előállításának képessége, és ezt szisztematikus - szabályok által vezérelt - egyszerűsítésekkel és helyettesítésekkel kompenzálják.

16.5 NYELV ÉS SZABÁLYOK

Bár ebben a fejezetben főként a fonológiai folyamatokra koncentráltunk, a nyelvészet más területei is bővelkednek a szabályvezérelt viselkedés példáiban. Ha már van némi tapasztalata a nyelvészettel kapcsolatban, vagy bevezető kurzust tervez, arra kérem, hogy gondolja át ismereteit - vagy közelítsen a témához - különös tekintettel arra, hogy ne csak a szabályvezérelt viselkedés bizonyítékait keresse, hanem olyan folyamatokat is, amelyek számíttással való magyarázata nehézségekbe ütközhet.

Az előző szakaszokban arra összpontosítottunk, hogy a nyelvi képességek számítógépes megvalósíthatósága mellett szóló bizonyítékokat azonosítsuk. A következő fejezetben visszatérünk a filozófiai anyaghoz, és problematizáljuk a nyelv előállításának és megértésének egy kulcsfontosságú - vitathatatlanul a leglényegesebb - aspektusát, nevezetesen a *jelentés* meghatározását.

17. FEJEZET

MEANING

Ez a fejezet hat fejezetnyi technikai anyag után visszatér a filozófiai anyaghoz.

Láttuk, hogyan lehet a számítógépeket úgy programozni, hogy stratégiailag összetett játékokat játsszanak, és ezt összehasonlítottuk azzal az intuitív felfogásunkkal, ahogyan az emberek ezeket a játékokat játsszák. Megnéztük a szakértői rendszereket mint a gépi érvelés példáját, és megvizsgáltuk a tipikus emberi teljesítményt bizonyos érvelési feladatokban annak meghatározásával összefüggésben, hogy a komputationalistáknak mekkora magyarázati terhet jelent e tipikus teljesítmény elszámolása.

Azt is megvizsgáltuk, hogyan alkalmazhatnánk formális rendszereket és keresési eljárásokat a nyelvi lehetőségben érintett mechanizmusok egyikének - a karakterláncok nyelvtani jellegének eldöntése - megkönnyítésére. Ezen túlmenően, a nyelvészetből olyan - főként a fonológiai folyamatokra vonatkozó - bizonyítékokat vontunk ki, amelyek alátámasztják azt az elképzelést, hogy a nyelvi viselkedés szabályvezérelt, és így számítással megvalósítható.

Ezután egy olyan gondolkísérletet fogunk megvizsgálni, amely a komputacionalizmust veszi célba, és azt próbálja megmutatni, hogy a mentális életnek van egy olyan döntő aspektusa, amellyel a komputacionalista nem tud elszámolni - az a tény, hogy mentális állapotaink *értelmesek*.

17.1 A KÍNAI SZOBA

Mentális állapotaink döntő és meghatározó jellemzője, hogy szemantikai tartalommal bírnak - hogy *jelentéssel bíró* állapotok. Minden megfelelő elméletnek képesnek kell lennie arra, hogy számot adjon a mentális állapotok szemantikai tartalmáról.

A számítás teljes mértékben szintaktikai folyamat. A formális rendszerek műveletei szintaktikailag meghatározott szimbólummanipulációk. Láttuk a formális rendszerek magyarázó effiktségét számos kognitív mechanizmus számbavételében. A döntő kérdés a

Jelenlegi célunk az, hogy a szemantikát a szintaktikai műveletekkel tudjuk-e számon kérni vagy sem.

A gondolat kísérletet, amelyet itt szeretnék bemutatni, eredetileg John Searle írta le, és annak megállapítására törekszik, hogy a szintaxis önmagában nem *elégséges* a szemantikához.

Képzelve el, hogy arra kéri, hogy kísérleti céllal több órán át végezzen el egy bizonyos feladatot. Bemutatkozik egy hatalmas teremben, amelyben több ezer polcnyi számozott könyv található. A szoba közepén van egy asztal, amelyen az egyik könyv van. Átlapozod a könyvet, és látod, hogy az nem tartalmaz mást, mint olyan szimbólumok átírási szabályait, amelyeket még soha nem láttál, minden egyes átírási szabály mellett egy számjegyzékkel.

Azt mondják, hogy egyedül maradsz a szobában, és ekkor egy papírdarabot, amelyen egy szimbólumsor van, átadnak az ajtóban lévő nyíláson. A feladatod az, hogy megkeresd a könyvben azt a szabályt, amelynek a bemeneti oldala pontosan ezt a szimbólumsorozatot tartalmazza, és a kimeneti szimbólumsorozatot másold ki a papírdarab másik oldalára, majd add vissza a nyíláson keresztül. Ezután meg kell keresned a polcokon azt a könyvet, amelynek száma megegyezik az imént követett szabály melletti számjegyekkel, és az asztalon lévő könyvet ezzel az új könyvvel kell kicserélned. Egyedül maradsz a szobában, és a dolgok pontosan a leírtak szerint zajlanak. Az ajtóban lévő nyíláson keresztül átadnak egy papírdarabot, te átkutatod az asztalon lévő könyvet, hogy megtaláld ezt a szálát egy szabály bemeneti oldalán, és kimásolod a szabály kimeneti szalagját, majd kicseréled az asztalon lévő könyvet a polcokról azzal a könyvvel, amelynek számát a szabály melletti jelölés adta meg. Egy másik papírdarabot egy új szimbólumsorral átvezetünk a résen a ajtó, és megismétli az eljárást.

Miután ezt több órán keresztül csináltad, azt mondták neked, hogy a szimbólumsorok valójában kínai írással írt mondatok voltak. A szobában lévő könyvek kódolják az összes lehetséges beszélgetést, amit több óra alatt kínaiul folytathatsz. Minden könyv egy beszélgetési állapotot képvisel, és ésszerű válaszokat ad a lehetséges bemenetekre.

Kiderül, hogy Ön több órán keresztül beszélgetett egy kínai anyanyelvűvel, és pusztán a könyvekbe kódolt átírási szabályok betartása alapján átment egy kínai Turing-teszten.

Nyilvánvaló azonban, hogy Ön ezáltal nem *érti a* kínai nyelvet. Először is, a kínai kérdésekre adott társalgási válaszok nem az ön hiedelmeivel és vágyaival, hanem a könyvekben kódolt önkényes válaszokkal voltak összhangban. Például az egyik kérdés lehetett az, hogy "Szereti az eperfagyálaltot?", és az Ön

a forgatókönyv szerinti válasz az volt, hogy "igen, nagyon finom", annak ellenére, hogy az eperfagylaltot ki nem állhatod. Vagy talán az egyik kérdés az volt, hogy "nem kezdesz éhes lenni?", és az előre megírt válaszd az volt, hogy "nem, egyelőre nem kérek, köszönöm", annak ellenére, hogy éhes vagy és kíváncsi vagy, mikor lesz ebéd.

Másrészt, a kínai nyelven való beszélgetés képessége nem terjed túl a kínai szobán. Ha egy kínai anyanyelvű beszélő átadna Önnek egy írott kínai kérdést, miután Ön elhagyta a szobát, a szimbólumsorok továbbra is értelmetlenek lennének az Ön számára. Csak a kínai szoba könyveiben található kódolt beszélgetési állapotok felhasználásával képes vagy a megértés látszatát kelteni, és átmenni a vonatkozó Turing-teszten.

A gondolatkísérletben leírt helyzet az, amelyben egy formális rendszer folyamatai - a szimbólumok formális szabályok szerinti átírása - megfelelnek egy Turing-tesztnek. A gondolatkísérlet által kiváltott intuíció azonban az, hogy bár a rendszer a megértés *látszatát* kelti - az ember meggyőzéséhez megfelelő módon -, a rendszer működéséből valami döntő fontosságú dolog hiányzik. Ezek önmagukban és önmagukban nem *jelentenek* semmit.

Más szóval, a kínai szoba szintaktikai műveletei, bár átmennek a Turing-teszten, nem rendelkeznek *szemantikával*.

17.2 SZINTAXIS ÉS SZEMANTIKA

A kínai szoba gondolatkísérlet alapján talán hajlamosak lennének ezt az érvet felhozni a komputacionalizmus ellen:

- P1 A szemantika megléte szükséges feltétele az elme meglétének.
 P2 A formális rendszerek szintaktikai műveletei nem elegendőek ahhoz, hogy szemantikával rendelkezzenek.

A formális rendszerek műveletei nem elegendőek az elme meglétéhez.

A komputacionalizmus hamis.

Az 1. premissza nem vitatható - egyértelmű, hogy a mentális állapotoknak szemantikai tartalma van. A 2. premissza azonban vitatható.

Kétféleképpen értelmezhetjük a második premisszát, és következésképpen a belőle következő időközi következtetést is. A gyengébb értelmezés az az állítás, amelyet a gondolatkísérlet engedélyez;

azonban egy sokkal erősebb értelmezésre hivatkoznak a komputualizmus állítólagos hamisságának levezetésében.

A második premissza gyenge értelmezése az, hogy létezik olyan formális rendszer, amelynek műveletei nem alkalmasak arra, hogy szemantikával rendelkezzenek. Az erősebb értelmezés az, hogy *nincs olyan* formális rendszer, amelynek műveletei alkalmasak a szemantika meglétéhez.

A kínai szoba gondolat kísérlet nem mutatja meg, hogy *nem* létezhethet olyan formális rendszer, amelynek műveletei alkalmasak a szemantika generálására. Következésképpen a fenti érvelés nem bizonyítja a komputualizmus hamisságát. Amit azonban a gondolat kísérlet megmutat, az a második premissza fentebb adott gyenge értelmezésénél valamivel erősebb.

A 2. feltétel gyenge értelmezése lényegében triviális. Adott, hogy rengeteg olyan formális rendszer létezik, amelynek műveletei nem felelnek meg az elme meglétének feltételeinek. A kínai szobás gondolat kísérlet érdekessége azonban az, hogy megmutatja, hogy létezik olyan formális rendszer, amelynek műveletei önmagukban is *megfelelnek a Turing-teszt teljesítéséhez*, mégis, intuitíve, a rendszerből teljesen hiányzik az értelem.

A kínai szoba gondolat kísérletét tehát úgy is értelmezhetnénk, mint a Turing-teszt efficacy-jának vádját. Végül is, ha valami át tud menni a teszten a megértés teljes hiánya ellenére, akkor úgy tűnik, hogy a teszt egyáltalán nem megbízható mutatója az elme jelenlétének. Mielőtt azonban ezt a következtetést levonnánk, el kell gondolkodnunk a gondolat kísérletben leírt rendszeren annak fényében, amit a formális rendszerekről és a természetes nyelvi feldolgozásról tudunk.

A gondolat kísérlet egy olyan rendszert ír le, amely bár *logikailag* lehetséges, *fizikailag* nem lehetséges. Ahhoz, hogy ezt a rendszert megvalósíthassuk, fel kellene állítanunk az összes lehetséges kínai beszélgetés generációs fáját, amely néhány óra leforgása alatt létrejöhet.

Tekintettel arra, hogy a lehetséges beszélgetési állapotok generációs fája lényegesen összetettebb lenne, mint a lehetséges sakkállapotok generációs fája, világosnak kell lennie, hogy még az első húsz lehetséges beszélgetésváltás teljes generációs fájának megalkotása számításilag egyszerűen nem kivitelezhető. Nyugodtan kijelenthetjük, hogy a gyakorlati számítási teljesítmény jövőbeli fejlődésétől függetlenül egyetlen számítógép sem lesz képes arra, hogy a kínai szoba által alkalmazott módszerrel átmenjen a Turing-teszten.

Általában egy gondolat kísérleti szituáció fizikai lehetősége ellen érvelni nem filozófiai munka, mivel általában gondolat kísérleteket

használunk arra, hogy *logikai* összefüggésekre vonatkozó állításokat teszteljünk.

A fizikalizmus például azt állítja, hogy a teljes fizikai leírás ugyanolyan *hiteles*, mint az elme teljes leírása. Ez a sufficiencia logikai állítás. Következésképpen, bár fizikailag lehetetlen, hogy létezzen egy olyan tudós, mint Mary, az 5. fejezetben leírt gondolat kísérlet mégis e sufficiencia ellen szól - amennyiben a gondolat kísérlet által előidézett intuíciónkat a gondolat kísérlet alapozza meg -, mivel a leírt helyzet logikailag lehetséges.

Ha a Turing-tesztre vonatkozó állítás hasonlóan *logikus* állítás lenne, akkor a kínai szoba gondolat kísérlete valóban ellene szólna. Emlékezzünk azonban arra, hogy az állítás nem az, hogy a Turing-teszt átvészélése az elme meglétéhez *szükséges*, hanem inkább az, hogy ha valami átmenne a Turing-teszten, akkor késznek kellene lennünk arra, hogy mentalitást tulajdonítsunk neki. A Turing-teszt egy *empirikus* teszt. Következésképpen, amikor a tesztnek mint az elme jelenlétének megbízható jelzőjének az efficacy-jára vonatkozó lehetséges ellenpéldákat mérlegeljük, a mérlegelésünket a Turing-tesztre kell korlátoznunk.

empirikusan lehetséges rendszerekhez.

A kínai szoba gondolat kísérlet tehát nem mutatja be egyértelműen a komputacionalizmus hamisságát, és nem vádolja a Turing-teszt helyességét. Valami nagyon fontosat azonban mégis megmutat.

A gondolat kísérlet szerintem azt mutatja, hogy a külső világtól elszigetelten semmilyen szintaktikai művelet nem alkalmas a szemantika létrehozására. Akár évekig is ülhetnék a kínai szobában és végezhetném ezt a műveletet - ha elég könyvet kapnék -, és intuitív módon úgy tűnik, hogy nem lesz módomban elkezdni megérteni az általam feldolgozott szimbólumok jelentését. Ennek az az oka, hogy műveleteimnek nincs megfelelő kapcsolata a külső világgal.

A kínai nyelv megértése *nem más, mint* annak megértése, hogy a nyelv - írott vagy beszélt - elemei hogyan kapcsolódnak a nyelvi rendszeren *kívüli* dolgokhoz. A nyelvek olyan rendszerek, amelyek jelentéseket kódolnak és közvetítenek. Ezeket a jelentéseket azonban nem a nyelvi létesítményen belülről származó mechanizmusok és bemenetek hozzák létre. Természetesen a nyelvi mechanizmusok részt vesznek a nyelvi entitások jelentéssel való felruházásában, de *szükségképpen* szükség van a külső világgal való megfelelő kapcsolatra is.

A kínai szoba gondolat kísérlet tanulsága az, hogy a szemantika fejlődéséhez szükség van a *megtestesült tapasztalatra*. Ahhoz, hogy mentális állapotaink jelentéssel bírjanak, előzetesen tapasztalunk kell a világban, amit érzékszerveink közvetítenek. Más szóval, a szemantika nem elszigetelten fejlődik, hanem ez a fejlődés az empirikus világgal kapcsolatos tapasztalatoktól függ.

A megtestesült tapasztalat szükségessége a szemantika fejlődéséhez önmagában nem a komputacionalizmus ellen szól. Csupán a komputualistákra háruló magyarázati terhet alakítja, megkövetelve tőlük, hogy a jelentést kölcsönző mechanizmusokról komputációs magyarázatot adjanak. Ez *többek* között magában foglalja az érzékelési adatok különböző típusú *mentális reprezentációkká* történő számításon átalakítását, amelyek aztán további ~~szerű~~ folyamatokban vesznek részt - például a természetes nyelvi kifejezések megértésében.

Itt különbséget kell tenni a szemantikai reprezentációk kialakulásának feltételei és azok között a feltételek között, amelyek mellett e reprezentációk tokenjei jelentéssel bírónak tekinthetők. Más szóval, elismerhetjük a megtestesült tapasztalat szükségességét a szemantikai reprezentációk fejlődéséhez, de aztán megfontolhatunk egy olyan gondolat kísérletet, amelyben a nyelvi produkciót és megértést elősegítő formális rendszer(ek) funkcionális megfelelőit egy teljesen fejlett anyanyelvi beszélőnél a kínai szobapéldát megközelítő módon valósítjuk meg, remélve, hogy tovább problematizáljuk egy ilyen rendszer műveleteinek értelmességét.

Emlékezzünk azonban a 14. fejezetből, hogy bár nincs okunk feltételezni, hogy ezek a nyelvi folyamatok nem számítási folyamatok, jó bizonyítékok támasztják alá azt az állítást, hogy a különböző mentális reprezentációk - köztük a szemantikai reprezentációk - összehangolt felhívása szükséges jellemzője a működésüknek.

Más szóval, nem arról van szó, hogy a lin- guisztikus megértés alapjául szolgáló folyamatok elszigetelt, moduláris folyamatok lennének, és a jelentés a feldolgozás utolsó szakaszában kapna értelmet. Ehelyett ezek a különböző folyamatok együttesen játszódnak le, a szemantikai reprezentációkhoz való folyamódásokkal, amelyek a fonológiai és szintaktikai folyamatok korlátozását és informálását szolgálják. Következésképpen - bár itt sem bizonyítékot, sem szilárd érveket nem hoztam - valószínűnek tűnik, hogy minden empirikusan lehetséges rendszer, amely alkalmas arra, hogy átmenjen a Turing-teszten, szükségszerűen tartalmaz értelmes szemantikai reprezentációkat.

Sokkal többet kell még mondani arról, hogy milyen feltételek mellett tekinthetők egy formális rendszer műveletei önmagukban értelmesnek (és nem csak értelmesként értelmezhetőnek). A következő fejezetben az értelmesség vizsgálatának fókuszát a *mentális reprezentáció* e fogalmának explicit vizsgálatára szeretném áthelyezni.

KÉPVISELET

Mentális állapotaink azáltal értelmesek, hogy dolgokról *szólnak*. Más szóval, az értelmes mentális állapotok *reprezentációs* állapotok - *reprezentálnak* vagy *képviselnek* dolgokat. Az előző fejezetekben utaltam a *mentális reprezentációkra*, például a fonémikus, szintaktikai és - ami döntő fontosságú - szemantikai reprezentációkra, amelyek megkönnyítik a nyelvi produkciót és megértést. Ebben a fejezetben a mentális reprezentációk szerkezetét és természetét szeretném röviden tárgyalni.

A reprezentáció meglehetősen kényes filozófiai téma. Nem célom, hogy ez a fejezet átfogó bevezetés legyen a mentális reprezentációval kapcsolatos különböző vitákba. Inkább arra szeretném használni a mentális reprezentáció természetéről szóló vitát, hogy különbséget tegyek a mesterséges intelligencia-kutatás két egymással versengő paradigmája között.

Céljaink szempontjából a legfontosabb különbség e két paradigma között - amelyeket *szimbolikus* és *konnekcionista* paradigmának fogok nevezni - a kutatók által a kognitív funkciók számításos reprodukálására alkalmazott módszerekben rejlik. A szimbolikus mesterséges intelligencia kutatója olyan szimbólumrendszereket fog alkalmazni, amelyeket már jól ismerünk, hiszen számos példát láttunk az előző fejezetekben. A konnekcionista mesterséges intelligencia kutatója ezzel szemben *mesterséges neurális hálózatokat* épít.

A mesterséges neurális hálózatokat - vagy más néven *kapcsolathálózatokat* - a 19. fejezetben vizsgáljuk meg részletesen. A következőkben szeretném világossá tenni, hogy a mesterséges intelligencia módszertanában ez a megkülönböztetés részben (bár nem teljesen) a mentális reprezentáció természetével kapcsolatos meggyőződésekből fakad.

18.1 SZÁNDÉKOSSÁG

Az intencionalitás a mentális állapotok reprezentatív természetének filozófiai szakkifejezése. Az *intencionális* állapotok azok, amelyek

valamiről szólnak, amelyek *reprezentálnak* valamit.

A "szándékosság" és a "szándékos" kifejezések nem tévesztendőek össze a "szándékozni" igével és annak rokon értelmű kifejezéseivel. Annak, hogy egy állapot ~~intenció~~ vagy sem, a filozófiai technikai értelemben semmi köze ahhoz, hogy egy ágens szándékában áll-e vagy sem. Inkább egy mentális állapot csak abban az esetben szándékos, ha valamiről *szól*. Az intencionalitás a mentális állapotok azon tulajdonsága, hogy azok a reprezentáció tárgyára (egy reprezentált dologra) irányulnak.

A mentális reprezentációk a szándékosság elsődleges hordozói. Mentális állapotaink szándékosak azért, hogy mentális reprezentációkat tartalmaznak alkotóelemként. Így például az a meggyőződés, hogy "a kutyám remek társ", *többek között* magában foglalja a "kutyám" jelképes mentális reprezentációját. A meggyőződés *a* kutyámról *szól*, mivel van egy konstitutív mentális reprezentációm, amely a kutyámról szól. A "kutyám" mentális reprezentációm a reprezentáció tárgyára - nevezetesen a kutyámra - irányul (arról).

Ezzel elérkeztünk a mentális reprezentációk egyik fontos jellemzőjéhez, amelyet itt szeretnék kiemelni. A mentális reprezentációk *kat-egorialisak*. A "kutya" mentális reprezentációm az összes és csakis a kutyákat választja ki. Más szóval, arra szolgál, hogy - mentális életemben - *kategorizálja* azokat a dolgokat, amelyeket kutyának tekintek, és megkülönböztesse őket azoktól a dolgoktól, amelyeket nem tekintek kutyának. Hasonlóképpen, a "barna" mentális reprezentációm minden és csakis azokat a dolgokat választja ki, amelyeket barna színűnek tartok, és a "szék" mentális reprezentációm minden és csakis azokat a dolgokat választja ki, amelyeket széknek tartok.

A mentális reprezentációk másik fontos jellemzője, amit szeretnék kiemelni, hogy *kompozíciós jellegűek*. A mentális reprezentációk összetettebb mentális reprezentációkba *állnak össze*. Ha például rendelkezem a "barna" és a "kutya" mentális reprezentációkkal, nincs szükségem többre ahhoz, hogy a "barna kutya" összetettebb mentális reprezentációját összeállítsam. Ez az összetettebb mentális reprezentáció kiválasztja az összes és csakis azokat a dolgokat, amelyeket barna kutyának tekintek.

A mentális reprezentációk e *kompozicionalitása* lehetővé teszi annak a magyarázatnak az egyik részét, hogy a mentális reprezentációk hogyan kapják meg szándékos tartalmukat. Az összetett mentális reprezentációk szándékosságukat a primitív mentális reprezentációktól öröklik, amelyekből összeállnak.

A döntő - és a legnehezebb - kérdés azonban az, hogy hogyan kapják meg a *primitív* mentális reprezentációk a szándékos tartalmukat. Hogyan jönnek létre atomi mentális reprezentációink a reprezentáció tárgyairól? Más szóval, milyen természetű a kapcsolat

a mentális reprezentációk és az általuk reprezentált tárgyak
(kategóriái) között?

18.2 KATEGÓRIÁK ÉS TARTALOM

A mentális reprezentációk és szándékos tárgyaik közötti kapcsolatról való beszámolás a mentális reprezentáció szemantikájának egy részét jelenti. A mentális reprezentáció szemantikájának számos elmélete létezik, de itt nem fogom kiegyensúlyozottan ismertetni a rendelkezésre álló elméleteket. Ehelyett csak *kétféle* elmélet legapróbb vázlatát szeretném adni.

Egyrészt vannak olyan elméletek, amelyek szerint a mentális ~~reprezentációk~~ alapvetően diszkrétnek. Másrészt vannak olyan elméletek, amelyek szerint a mentális reprezentációk alapvetően inter- láltak.

Az első típusú elméletek jól illeszkednek a mesterséges intelligencia kutatásának szimbolikus paradigmájához. Az ilyen típusú elmélet szerint a mentális reprezentációk szimbólumok.

A második típusú elméletek jól illeszkednek a mesterséges intelligencia kutatásában a konneccionista para- digmához. Az ilyen típusú elmélet szerint a mentális reprezentációk elosztott minták.

Egyrészt a szimbolikus reprezentáció, másrészt az elosztott reprezentáció melletti elkötelezettség egy sor járulékos elkötelezettséget hoz magával. Ezek közé tartoznak azok az elkötelezettségek, amelyek a reprezentációk szándékos tartalmának felruházására szolgáló mechanizmusokra, a reprezentált kategóriák természetére és szerkezetére, valamint a mentális reprezentációk kölcsönhatásának módjára vonatkoznak.

A szimbolikus reprezentáció hívei a reprezentációkat több szempontból is alapvetően diszkrétnek tekintik. Azt a mechanizmust, amely révén a szimbólumokat tartalmukkal ruházzák fel, úgy értelmezik, mint valamiféle közvetlen kapcsolatot a szimbólum tokenjei és a reprezentáció tárgyai között. Lényeges, hogy ez a mechanizmus olyan, hogy egy szimbólum tartalma semmilyen módon nem függ más szimbólumok tartalmától. Minden egyes szimbólumot külön-külön ruházunk fel a maga szándékolt tartalmával.

Továbbá a szimbolikus reprezentációkat úgy értelmezzük, hogy más reprezentációkkal való kölcsönhatásaikban diszkrétnek maradnak. A mentális reprezentáció kompozicionalitását egyszerű szín-taktikai összekapcsolásként értelmezik. Amikor a szimbólumok összetételükkel összetettebb reprezentációkat alkotnak, minden szimbólum mindig ugyanazt a tartalmat hozza a komplexumhoz, amelyben részt vesz. Más szóval, a szimbólumok tartalmát kontextusfüggetlennek tekintjük.

A szimbólumok további jellemzője, hogy jelenlétük bináris - egy szimbólum token vagy jelen van, vagy nincs. Ha egy szimbólumjelző jelen van, akkor teljesen jelen van, ha pedig nincs jelen, akkor teljesen hiányzik. A szimbólumok, ha úgy tetszik, vagy

be vannak kapcsolva, vagy ki vannak kapcsolva, a kettő között nincs lehetőség. Ez a

A szimbolikus reprezentáció bináris jellege az általuk reprezentált kategóriák természetével kapcsolatos következményekkel jár.

Ha a mentális reprezentációk szimbolikusak, akkor az általuk reprezentált kategóriáknak éles határokat és belső struktúrát nem ismerhetnek el. Más szóval, ha a mentális reprezentációk szimbolikusak, akkor az általuk reprezentált kategóriák olyanok, mint a dobozok - a tárgyak vagy benne vannak a kategóriában, vagy nem, és nincsenek jobb vagy rosszabb esetek a kategóriához tartozásra.

A szimbolikus reprezentáció hívei, összefoglalva, a reprezentációk tartalmát adó mechanizmust diszkrétnek, az általuk reprezentált kategóriákat binárisnak és strukturálatlannak, a mentális reprezentáció kom- pozícióját pedig kontextusérzékeny szintaktikai összekapcsolásnak tekintik.

Az elosztott reprezentáció hívei viszont a mentális reprezentáció szemantikájának minden egyes elemét differen értelmezik.

18.3 SZIMBÓLUMOK ÉS MINTÁK

Azok az elméletalkotók, akik a mentális reprezentációkat mint széttagolt mintákat tartják számon, a reprezentációkat alapvetően egymáshoz kapcsolódónak tekintik. A mentális reprezentáció szemantikája, amelyet az ilyen teoretikusok képviselnek, olyan, hogy azt a mechanizmust, amellyel egy reprezentáció tartalmat kap, alapvetően más reprezentációkkal való kapcsolatok közvetítik.

Ezt a közvetítést többféleképpen lehet konkretizálni, de az egyes elméletek mechanikája itt nem tartozik ránk. Ami a mi céljaink szempontjából fontos, az a mentális reprezentációk összefüggő természete melletti elkötelezettség, ami szöges ellentétben áll a szimbolikus reprezentáció híveinek nézetével. Az a mód, ahogyan egy mentális reprezentációnak szándékos tartalma lesz, a reprezentáció elosztott elmélete szerint alapvetően összefügg azzal, ahogyan más mentális reprezentációknak szándékos tartalma lesz.

Az elosztott mentális reprezentációk kompozícióját is valamivel összetettebbnek tekintjük, mint a szintaktikai összekapcsolást. A megismerés szisztematikusságának biztosításához elengedhetetlenül fontos, hogy a mentális reprezentációk kompozíciójának olyan leírása legyen, amely alkalmas a megismerés szisztematikusságának biztosítására, mivel általában úgy tartják, hogy ez a lin- guisztikus lehetőség produktivitását magyarázza.

Azok, akik a mentális reprezentáció elosztottként való felfogását támogatják, a reprezentációk összetételét úgy értelmezik, hogy az erősen

az aktivációs minták összetett kölcsönhatása egy elosztott hálózatban. Hogy ez pontosan mit jelent, az a következő fejezetben válik majd világosabbá, amikor a mesterséges neurális hálózatokról lesz szó. Jelen célokra elegendő annak megértése, hogy az elosztott reprezentációk összetételének módja kontextuálisan modulált. Más szóval, az a tartalom, amelyet egy adott reprezentáció a komplex reprezentációba visz, amelyben részt vesz, a komplexben szintén részt vevő többi adott reprezentációtól függő módon változik.

Az elosztott reprezentációk további jellemzője, hogy részben tokenizálhatók. Mivel a token-reprezentációkat úgy vesszük, hogy az aktiválási mintázatok széles körben eloszlanak a csomópontok összekapcsolt hálózatában, ezek a mintázatok részben aktiválhatók. Ismétlem, hogy ez pontosan mit jelent, a következő fejezetben válik majd világosabbá. A fontos pont itt azoknak a következményeknek a megbecsülése, amelyekkel a mentális reprezentációk részleges tokenizálásának lehetősége jár a reprezentált kategóriák természetét illetően.

Ha a mentális reprezentációk olyan elosztott mintázatok, amelyek részben tokenizálhatók, akkor az általuk képviselt kategóriák pontatlan határokat és belső struktúrát is megengedhetnek. Továbbá, ha a reprezentációk tartalma kontextuálisan modulált, akkor a kategória kiterjesztése kontextuálisan érzékeny lesz.

Más szóval, ha a reprezentációk elosztottak és kontextuálisan moduláltak, akkor az általuk reprezentált kategóriák olyanok, hogy a tagságnak lehetnek határesetei, a határok kontextusfüggően eltolódhatnak, és létezhetnek jobb és rosszabb eseteket megengedő, fokozatos tagságok.

Összefoglalva, az elosztott reprezentáció hívei a reprezentációk tartalmát átadó mechanizmust alapvetően más reprezentációkkal való kapcsolatok által közvetítettnek, az általuk reprezentált kategóriákat kontextuálisan érzékenynek - ami pontatlan és változó határokat és belső struktúrát tesz lehetővé -, a mentális reprezentáció összetételét pedig egy erősen összekapcsolt hálózatban lévő aktiválási módok komplex, kontextuálisan modulált kölcsönhatásának tekintik.

18.4 KOGNITÍV ARCHITEKTÚRA

Ebben a fejezetben eddig a mentális reprezentáció két különböző nézetét tárgyaltam, és ezt a megkülönböztetést használtam fel a mesterséges intelligencia kutatásában egymással versengő szimbolikus és konnekcionista paradigmák megértéséhez.

Ezek a mentális reprezentációra vonatkozó differáns nézetek központi jelentőségűek a két paradigma megkülönböztetésében, de

nem merítik ki a közöttük lévő különbségeket. A konnekcionisták a *kognitív architektúrára* vonatkozó nézetek tekintetében is eltérnek szimbolikus társaiktól.

A *kognitív architektúra* kifejezés egy kognitív ágens információfeldolgozó rendszereinek szerkezetére és természetére utal. Más szóval, a kifejezés a kogníciót elősegítő számítási hardver szervezési és implementációs jellemzőire utal.

A mesterséges intelligencia kutatásának szimbolikus hagyománya az emberi elme kognitív architektúráját *fizikai szimbólumrendszernek* tekinti. A konnekcionisták ezzel szemben az emberi kognitív architektúrát *konnekcionista hálózatokban* látják, amelyek lehetővé teszik a *párhuzamos, elosztott feldolgozást*.

Az előző fejezetekben számos példát láttunk arra, hogyan valósíthatunk meg kognitív funkciókat szimbólumrendszerekkel. A konnekcionista hálózatok, mint a következő fejezetben látni fogjuk, különösen alkalmasak olyan funkciók megvalósítására, amelyeket köztudottan nehéz szimbólumrendszer-architektúrában megvalósítani.

Amennyiben a konnekcionista architektúra könnyen alkalmazható olyan funkciók megvalósítására, amelyeket a megismerés fontos alkotóelemének tekintünk, és amelyek megvalósítása szimbólumrendszerekkel problémásnak bizonyul, legalább egy okunk van arra, hogy a konnekcionista megközelítést előnyben részesítsük a mesterséges intelligencia szimbolikus megközelítésével szemben.

A következő fejezetet annak szenteljük, hogy tisztázzuk azokat a fogalmakat, amelyekről eddig csak említés történt, kevés magyarázattal. E fogalmak magyarázata és a konnekcionista hálózatok működésének számos példával való szemléltetése után visszatérünk a szimbolikus és a konnekcionista paradigma közötti kapcsolat további tárgyalására.

19. FEJEZET

MESTERSÉGES NEURÁLIS HÁLÓZATOK

A konnekcionista paradigma a mesterséges intelligencia kutatásában a huszadik század utolsó két évtizedében vált ismertté. A mesterséges neurális hálózatokról bebizonyosodott, hogy igen jól modelleznek bizonyos kognitív jelenségeket, amelyek szimbolikus számítási architektúrával való megvalósítása problémás volt.

A mesterséges neurális hálózatok működését úgy tervezték, hogy az agy idegi áramköreit utánozzák - gyakran nevezik őket "agyi stílusú" feldolgozás megvalósításának. Ezért segíthet a fejezet megértésében, ha először a 4. fejezetben a neuronok működésének tárgyalását tekintjük át.

Ebben a fejezetben a mesterséges neurális hálózatok működésének és a kognitív funkciók módosításában való hasznosságuknak a megalapozott megértése következik. Azzal kezdjük, hogy ismertetjük az alapvető konnektoros architektúrát, és elmagyarázzuk, hogyan különbözik ez a szimbolikus számítási architektúrától.

19.1 KONNEKCIONISTA ARCHITEKTÚRA

A klasszikus szimbolikus számítási architektúra - amelyet a 7-9. fejezetben hosszasan ismertettünk, és amelyre azóta számos példát láttunk - a következő alapvető jellemzőket ismeri el.

Először is, az architektúrában csak egy processzor van - egy központi feldolgozó egység (CPU), amely a programutasításokat dolgozza fel. Másodsor, a CPU ezeket az utasításokat sorozatosan - egymás után - hajtja végre. Harmadsor, a CPU lokalizált regisztertartalmakat címez meg, és azokon dolgozik. A konnekcionista

architektúra ezzel szemben döntően különbözik a fenti jellemzők mindegyikét tekintve. A konnekcionista hálózatok (jellemzően nagyszámú) egyszerű feldolgozóegységekből (csomópontokból) állnak, amelyek nem sorosan, hanem párhuzamosan működnek. A tartalom a konnekcionista hálózatokban nem helyi és címezhető, hanem elosztott.

számos csomóponton keresztül, és a kapcsolatok mintájaként kódolva.

A mesterséges neurális hálózatok alapelemei egyszerű feldolgozó egységek, amelyeket úgy terveztek, hogy az egyes neuronok működését utánozzák. Ezek az egységek funkcionálisan rétegekbe szerveződnek - lesz egy csomópontokból álló bemeneti réteg és egy csomópontokból álló kimeneti réteg. Általában van egy "rejtett" csomóponti réteg is - ezek nem bemeneti és nem kimeneti egységek, hanem a rétegek közötti közvetítésre szolgálnak.

Amint azt már bizonyára Ön is megállapította, a csomópontok egymással kapcsolatban állnak. Az, hogy pontosan hogyan vannak összekapcsolva, különböző építészeti variációkat határoz meg, amelyekkel itt nem kell foglalkoznunk. Az érdekes bonyolultságú hálózatokban minden egyes csomópont nagyszámú más csomóponttal lesz összekapcsolva - ahogyan az egyes neuronok is nagyszámú más neuronhoz kapcsolódnak. A konnekcionista architektúra legegyszerűbb típusa (vagy a legbonyolultabb, attól függően, hogyan nézzük) olyan, hogy minden csomópont a hálózat minden más csomópontjával kapcsolatban áll.

A mesterséges neurális hálózatokban az információfeldolgozás az *aktivációnak* a hálózaton belüli kapcsolatok mentén történő terjedésével valósul meg. A hálózat minden egyes csomópontja rendelkezik egy aktivációs szinttel, amelyet a hozzá kapcsolódó más csomópontoktól kapott aktiváció befolyásol.

Az aktiválással kapcsolatban néhány egyszerűsítő feltételezést teszünk. Először is, feltételezzük, hogy minden egyes időlépésben egy csomópont aktiválódását teljes mértékben a bejövő (különböző) kapcsolatai mentén kapott aktiválódás határozza meg (ahelyett, hogy egy bonyolultabb függvényt vennénk figyelembe, amely figyelembe veszi a csomópont előző időlépésben kapott aktiválódási szintjét is).

A csomópontok közötti kapcsolatok lehetnek *gerjesztőek* vagy *gátlóak*, és ezt úgy reprezentáljuk, hogy minden kapcsolathoz egy *súlyt* - pozitív vagy negatív számértéket - rendelünk. A gerjesztő kapcsolatok - amelyek pozitív súlyozásúak - *gerjesztik* (növelik a csomópont aktiváltsági szintjét) azt a csomópontot, amelyhez kapcsolódnak. A gátló kapcsolatok - amelyek negatív súlyúak - *gátolják* (csökkentik az aktivációs szintet) azt a csomópontot, amelyhez kapcsolódnak.

A hálózat minden egyes csomópontja, mint emlékszik, egy egyszerű feldolgozóegység. Ezek a csomópontok két funkciót valósítanak meg - egy *aktiválási* függvényt és egy *átviteli* függvényt.

Az *aktiválási* függvény határozza meg, hogy egy csomópont az adott időlépéskori aktiváltsági szintje alapján *tüzel-e* vagy sem. Mi csak a legegyszerűbb aktiválási függvényt fogjuk megvizsgálni - a küszöbfüggvényt. A küszöbérték-aktiválási függvénnyel rendelkező

csomópontok akkor fognak tüzelni, ha az adott időlépésben az aktiválási szintjük meghaladja a következőhöz rendelt küszöbértéket

a csomópont. Ha egy csomópont tüzel, akkor az aktiválást minden egyes kimenő (efferent) kapcsolatán keresztül továbbítja más csomópontokhoz, egyébként az adott csomóponton keresztül nem terjed aktiválás.

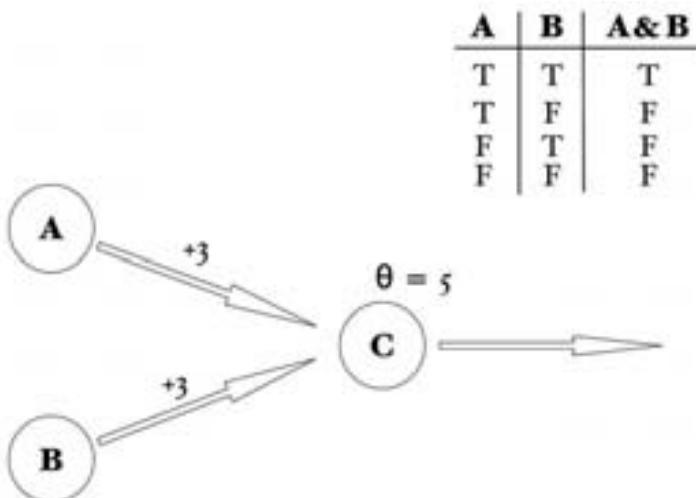
Az *átviteli* függvény határozza meg, hogy egy csomópont hogyan frissíti aktivációs szintjét a különböző összeköttetései mentén kapott aktiválás alapján. Ismét csak a legegyszerűbb átviteli függvényt - a *súlyozott* összegfüggvényt - fogjuk megvizsgálni. Egy csomópont aktivációs szintjének meghatározásához egy súlyozott összegű átviteli függvénnyel egyszerűen az afferent kapcsolati súlyok értékeinek összegét vesszük.

19.2 EGYSZERŰ MESTERSÉGES NEURÁLIS HÁLÓZATOK

Nézzünk meg néhány alapvető példát, hogy szemléltetni tudjuk ezeket a műveleteket. Az egyszerűség kedvéért egész számokat fogok használni a kapcsolati súlyok és a küszöbértékek esetében. A 19.1. ábra a legegyszerűbb mesterséges neurális hálózatot ábrázolja, amely valami érdekeset csinál.

Ez a hálózat két bemeneti csomóponttal (A és B) és egy kimeneti csomóponttal (C) rendelkezik. Minket az érdekel, hogy a kimeneti csomópont tüzel-e vagy sem (bár annak efferent kapcsolata nem afferent egyik másik csomóponthoz sem). A bemeneti csomópontokat valamiféle detektoroknak képzelhetjük el. Ezek akkor lépnek működésbe, ha valamilyen környezeti feltétel teljesül - talán ha egy lámpa ég, vagy ha egy kapcsoló egy bizonyos állásban van.

A hálózat két kapcsolata mindkettő gerjesztő és egyenlő súlyú. Ha A tüzel, akkor C-t gerjeszti, ha pedig B tüzel, akkor C-t gerjeszti.



19.1 ábra Számítás ÉS.

A C küszöbérték 0 olyan, hogy csak akkor lép működésbe, ha A és B is működésbe lép. Ha A egyedül tüzel, akkor a C aktiválási értéke 3 lesz, ami az 5 küszöbérték alatt van. Hasonlóképpen, ha B egyedül tüzel. Ha azonban mindkettőn tüzelnek, akkor a súlyozott összegű átviteli függvény azt mondja, hogy a C aktiválási értéke a különböző kapcsolati súlyok értékeinek összege lesz, ami ebben az esetben 6 lesz. Ez a 6-os aktiválási érték magasabb, mint a C-hez rendelt 5-es küszöbérték, így a C a küszöbérték aktiválási függvényének megfelelően fog tüzelni.

Ez a hálózat *logikai kapuként* szolgál. Kiszámítja a bináris logikai igazságfüggvényt. A kimeneti egység tüzel, ha A és B is tüzel.

19.1. gyakorlat

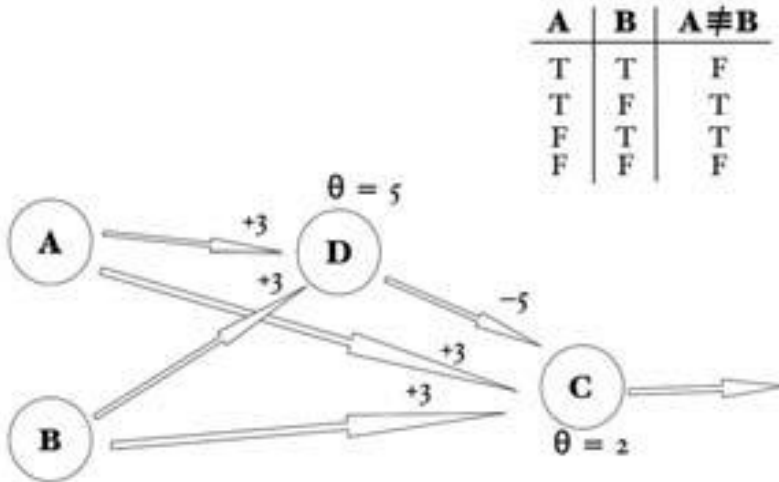
- (a) Hogyan módosíthatnánk az ábrán látható hálózatot?
19.1 úgy, hogy kiszámítja a *diszjunkció* bináris logikai függvényét - azaz úgy, hogy a kimeneti csomópont iff vagy A vagy B (vagy mindkettő) tüzel.
- (b) Tervezzon egy olyan hálózatot két bemeneti csomóponttal és egy kimeneti csomóponttal, hogy a kimeneti csomópont akkor lépjen működésbe, ha valamelyik bemeneti csomópont lép működésbe, de akkor *ne*, ha mindkettő lép működésbe.

Ha a 19.1(b) feladatot sikeresen megoldotta, akkor olyan hálózatot tervezett, amely kiszámítja a *kizárólagos diszjunkció* bináris logikai igazságfüggvényét. Ez nem egy egyszerű feladat, mivel két olyan dolgot kell tennünk, ami újdonság számunkra. Az egyik a gátló súlyok hozzárendelése; a másik pedig az, hogy a bemeneti és a kimeneti egységek közé egy rejtett egységet adjunk a hálózathoz. A megoldás a 19.2. ábrán látható.

Ha egy pillanatra figyelmen kívül hagyjuk a 19.2. ábra D csomópontját, akkor megkapjuk a 19.1(a) feladat megoldását. Mindössze annyit kellett tennünk, hogy a C küszöbértékét a különböző kapcsolati súlyok valamelyikénél alacsonyabb értékre csökkentettük. (Alternatív megoldásként mindkét kapcsolati súlyt a küszöbérték feletti értékre emelhetjük volna). Ahhoz azonban, hogy megakadályozzuk, hogy C akkor tüzeljen, amikor A és B is tüzel, hozzá kell adnunk a D csomópontot.

A D csomópont tüzelni fog, ha A és B is tüzel, és gátolja a C csomópont aktiválását, hogy megakadályozza a tüzelést. Ha A egyedül tüzel, akkor a D aktiválási értéke a küszöbérték alatt lesz, és nem fog tüzelni, de a C aktiválási értéke a küszöbérték felett lesz, és tüzelni fog. Hasonlóképpen, ha B egyedül tüzel. Ha azonban A és B

is tüzel, akkor D *is* tüzel, mivel aktiválódási értéke a küszöbérték felett lesz. A C aktiválási értéke a különböző kapcsolati súlyok összege lesz ($3 + 3 + -5 = 1$).



19.2 ábra XOR számítás.

amely a küszöbértéke alatt van, így mivel a D által gátolt, nem fog tüzelni. Összefoglalva, C akkor fog tüzelni, ha A vagy B tüzel, de akkor *nem*, ha mindkettő tüzel, *quod erat demonstrandum*.

19.3 BESZÉDSZINTETIZÁLÁS

Ebben a részben egy mesterséges neurális hálózatot tervezünk, amely angol beszédszintetizátorként működik. Ez egy szép példa arra, hogy a konnektív hálózatok milyen kontextusérzékeny feldolgozási feladatokban jeleskednek.

Az angol ortográfia nem fonémikus - nem lehet szabályosan leképezni az angol fonémákra. Ellentétben például a japán nyelvvel, amely úgy működik, hogy egy adott grapéma kiejtése kontextuálisan változatlan, egy adott angol grapéma kiejtése az ortográfiai kontextustól függ. Más szóval, az angol nyelvet nem úgy ejtik, ahogyan írják, hogy lazán fogalmazzak. Sokkal inkább az a hang, amelyet egy adott betű jelöl, kontextusfüggő - egy és ugyanaz a betű differens hangokat jelölhet, és ugyanazt a hangot különböző betűkkel lehet ábrázolni, a helyesírási kontextustól függően.

Következésképpen egy angol nyelvű beszédszintetizátor tervezése *többek között* magában foglalja az ortográfiai reprezentációk fonémás reprezentációvá alakításának kontextusfüggő feldolgozási feladatát. Elkezdjük egy mesterséges neurális hálózat megépítését az ortográfiai bemenetnek fonémás kimenetté való átalakításának megvalósítására.

Ha azt akarjuk, hogy a hálózatunk kontextusérzékeny legyen, akkor nyilvánvalóan szükségünk van a bemeneti réteg kontextusára, hogy érzékenyek legyünk rá. Ezt úgy fogjuk elérni, hogy a bemeneti réteget öt csomópontcsoportba szervezzük. Mindegyik pool huszonzét csomópontot tartalmaz, amelyek a betűdetektorok teljes készletét képviselik - egyet az ábécé minden egyes betűjéhez, valamint egyet a szóközhöz (az egyszerűség kedvéért az írásjeleket itt figyelmen kívül hagyjuk).

Ezeket a bemeneti poolokat úgy kell megszervezni, hogy mindegyik pool a bemeneti szöveges karakterlánc egy-egy betűhelyére irányuljon. Minden egyes időlépéskor egy pool egy célbetű pozíciójára irányul. A másik négy pool úgy lesz megszervezve, hogy egy-egy pool a célpozíció két oldalán lévő két betűhelyre irányuljon. Ez lehetővé teszi a hálózat számára, hogy a környező ortográfiai kontextus (a célbetű két oldalán lévő két betű) alapján kontextuálisan meghatározza, hogy egy adott betű melyik fonémát jelöli. Az első időlépésnél a szövegsorozat első betűje kerül a célpozícióba. Minden ezt követő időlépésnél a szövegsorozatot úgy haladjuk előre, hogy a sorban a következő betű kerüljön a célpozícióba. A beszédszintetizáló hálózatunk kimeneti rétege minden egyes fonémához egy kimeneti csomópontból áll, így ha az ausztrál angol nyelvet vesszük figyelembe, akkor negyvennégy kimeneti csomópont lesz. Az egyszerűség kedvéért itt csak egy fonémát fogunk figyelembe venni, amelynek kiejtése az angol nyelvjárásokban változatlan: /s/ - a "csók" szó véghangja.

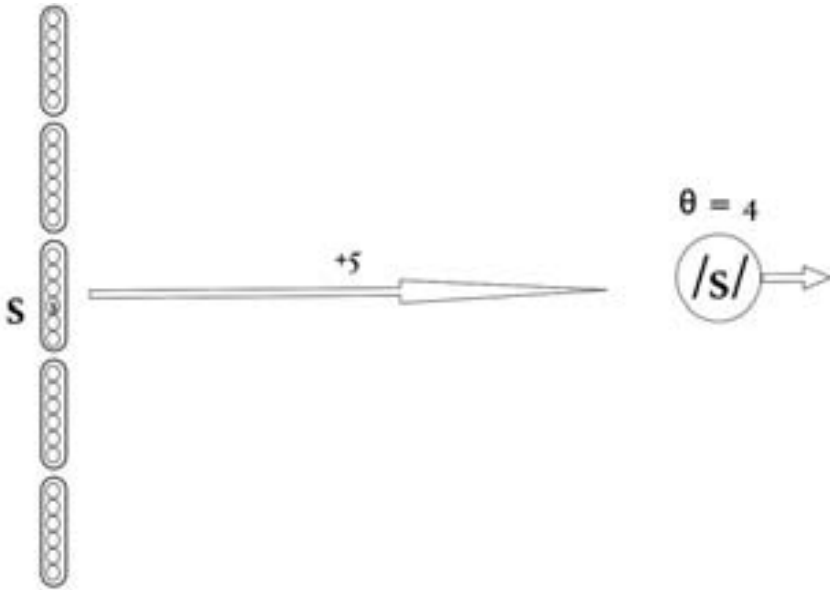
és 'ez'.

Annyi rejtett egységgel segíthetünk magunknak, amennyire szükségünk van ahhoz, hogy a bemeneteket helyesen illesszük a kimenetekhez. Olyan küszöbértékeket és kapcsolati súlyokat fogunk beállítani, hogy a hálózatunk helyesen döntse el, hogy az /s/-t ki kell-e ejteni a következő teszt szavak esetében: this, gas, wish, shy, kiss, passive, asia, asian, asiatic, is, as, ice, justice, service.

Az első dolog, amit meg kell tennie, az a szabványos tok elhelyezése. Amikor az "s" betűt látjuk, általában az /s/ fonémát kell előállítani. Ezért az első dolog, amit teszünk, hogy a célbetű pozíciójához tartozó bemeneti poolban lévő 's' detektort közvetlenül a /s/ fonémát reprezentáló kimeneti egységhez kapcsoljuk, úgy, hogy ha az 's' betűt a célhelyzetben észleljük, akkor az /s/ egység tüzel, hacsak másképp nem gátoljuk (lásd a 19.3. ábrát).

Hálózatunk most már helyesen fog működni a tesztkészletünk első két szava - "ez" és "gáz" - tekintetében. Amikor az "s" mindkét szóban eléri a célpozíciót, a /s/ egység ki fog löni, ahogyan azt kell.

Számos olyan szó van azonban, amelyben az "s" betű nem az /s/ fonémát jelenti. A születőben lévő beszédszintetizáló hálózatunk



19.3 ábra A standard eset.

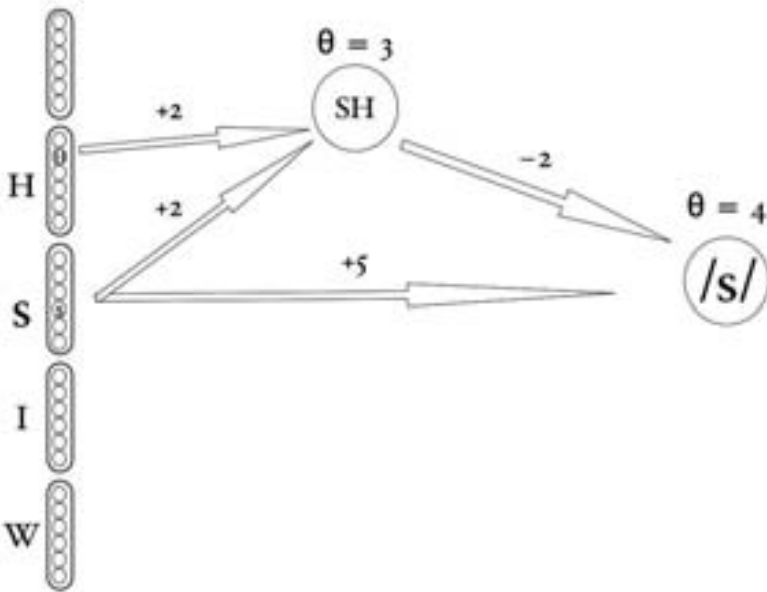
jelenleg helytelen meghatározásokat fog tenni a tesztlistánkon szereplő többi szóval kapcsolatban. A következő feladatunk tehát az, hogy olyan rejtett egységeket tervezzünk, amelyek felismerik azokat az összefüggéseket, amelyekben az "s" betű megjelenik, de az /s/ fonémát nem kellene előállítani, és ezeket a rejtett egységeket arra használjuk, hogy ennek megfelelően gátolják a kimeneti egység aktiválását.

Azt szeretnénk, ha a rejtett rétegben reprezentált kontextusok a lehető legáltalánosabbak lennének, hogy a lehető legtöbb eset figyelembe lehessen venni. Az angol nyelvben - néhány tulajdonnév és összetett szó kivételével - szinte mindig az a helyzet, hogy ha az "s" betűt egy "h" betű követi, akkor azt nem /s/-ként ejtjük. Az első rejtett egység, amelyet a hálózathoz hozzáadunk, éppen az ilyen kontextusokat fogja érzékelni, és gátolja a kimeneti egységet (lásd a 19.4. ábrát).

Most, amikor a hálózatunk elé kerül a tesztkészletben szereplő két következő szó - "wish" vagy "shy" - valamelyike, a hozzáadott rejtett egység gátolni fogja a /s/ egységet, hogy az ne lépjen működésbe, ahogyan szükséges. Amint az világos, a hálózat mostantól minden olyan ortográfiai kontextust, amelyben az "s"-t "h" követi, olyan kontextusnak tekint, amelyben az /s/ nem keletkezik.

Hasonló szabályosságot mutat az a szöveggörnyezet is, amelyben az "s"-t egy másik "s" követi, mint például a "csók" és a "passzív" szavak. Ezekben az esetekben az /s/ fonéma csak egyszer keletkezik.

A jelenlegi helyzetben a hálózatunk azt fogja megállapítani, hogy az /s/-t kétszer kell kiejteni, mivel a kimeneti egység akkor fog tüzelni, amikor minden egyes "s" a célpozícióban van.



19.4 ábra Kontextusérzékenység.

Következésképpen egy újabb rejtett egységet kell hozzáadnunk a kettős "s" kontextusok befogadásához (lásd a 19.5. ábrát).

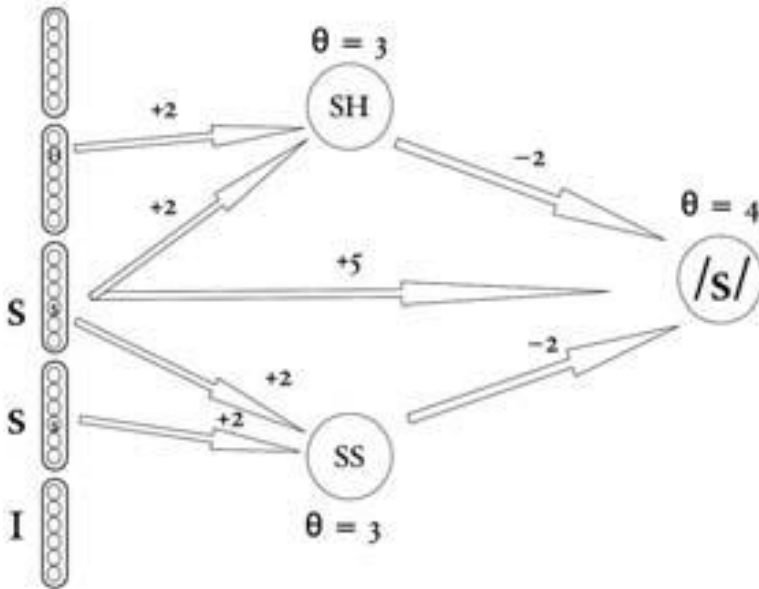
Megjegyezzük, hogy a dupla "s" problémára egy kissé diffparátusú megoldást is alkalmazhattunk volna. Nem számít, hogy melyik "s" betű esetében gátoljuk meg a /s/ fonéma produkcióját, amíg csak egyszer produkáljuk. Így ugyanúgy meggátolhattuk volna a kimeneti csomópont tüzelését egy olyan "s" esetén is, amelyet egy következő "s" követ, nem pedig egy megelőző "s". A tesztkészletünk következő három szava - "asia", "asian" és "asiatic" - mindegyiket egyetlen rejtett egység hozzáadásával tudjuk kezelni. Az angol nyelvben nincs olyan szó az "asia" betűkombinációval, amelyben az "s"-t /s/-ként ejtjük. Ezért hozzáadhatunk egy rejtett csomópontot a hálózathoz, amely érzékeli az "asia" kontextust, és gátolja az "asia" kapcsoló tüzelését.

kimeneti egység.

19.2. gyakorlat

A hálózat bővítése egy rejtett egység hozzáadásával, valamint a súlyok és küszöbértékek beállításával az "asia" kontextus észlelésére és a kimeneti egység tüzelésének gátlására.

A tesztkészletben szereplő két következő szó - az "is" és az "as" - mindegyikének befogadása több specifikumot igényel. Bár azt akarjuk, hogy a kontextusunk



19.5 ábra Kontextusérzékenység.

detektorok a lehető legáltalánosabbak legyenek, ez a két eset nem engedi meg a kontextuális általánosítást. Megkísérhetnénk például, hogy a kimeneti egység tüzelését minden olyan kontextusban gátoljuk, amelyben az "is" a szó végén jelenik meg. Ez ugyan az "ő"-t is befogadná, de a hálózat ezt követően hibás meghatározást tenne, amikor a "ez" szóval találkozánk. Hasonlóképpen az "as", "has" és "gas" esetében is.

Következésképpen olyan rejtett egységeket kell hozzáadnunk, amelyek csak az "is" és az "as" szavakat érzékelik. Itt válik nyilvánvalóvá annak hasznossága, hogy minden bemeneti medencében van egy csomópont a szóköz felismerésére. Az "is" szó észlelése a "_is_" kontextus észlelését jelenti, ahol az "_" egy szóközt jelöl.

19.3. gyakorlat

A hálózat bővítése rejtett egységek hozzáadásával, valamint súlyok és küszöbértékek beállításával a "_mint_" és "_van_" kontextusok észlelése és a kimeneti egység tüzelésének gátlása érdekében.

Amellett, hogy számos olyan kontextus létezik az "s" betűvel, amelyben a /s/ fonéma nem keletkezhet, számos olyan orto-grafikus kontextus is létezik, amelyben a /s/ fonémának keletkeznie kellene, és amelyek nem tartalmazznak "s"-t. A tesztkészletünk utolsó két szava

200
éppen ilyen eset.

Majdnem minden angol szó, amely az "ice" betűkombinációt tartalmazza, olyan, hogy a "c"-t /s/-ként ejtik. A legegyszerűbb módja a tesztkészletünkben megmaradt szavak befogadásának az, hogy hozzáadunk egy rejtett csomópontot, amely érzékeli ezt a kontextust, és úgy gerjeszti a kimeneti egységet, hogy az szükség szerint tüzelni fog.

19.4. gyakorlat

A hálózatot egy rejtett egység hozzáadásával, valamint a súlyok és küszöbértékek beállításával bővítheti a "jég" kontextushoz.

19.5. gyakorlat

Bővítse tovább a hálózatot, hogy a következő szavak is helyet kapjanak: precíz, visszahúzódni, érzékelni, kapni, kicsapni, vétel, recept.

19.6. gyakorlat (Kihívás)

- (a) Próbáljon meg a következő szavak közül minél többet beilleszteni anélkül, hogy a hálózat hibás döntéseket hozna az eredeti tesztkészletünkben vagy a 19.5. feladatban szereplő kibővített készletben szereplő szavak bármelyikével kapcsolatban: ease, lease, please, peace, grease, guise, reprise, practise, practice, his, has, mission, passion.
- (b) Mi akadályozza meg, hogy mindezeket a szavakat befogadjuk? Hogyan bővíthetnénk a hálózati architektúránkat, hogy javítsunk ezen?

19.4 TANULÁS

Most már pontosan érzékelnie kell, hogy milyen összetett feldolgozási feladat az angol helyesírás fonémává alakítása. Csak egy fonémával és a releváns eseteknek csak egy apró töredékével foglalkoztunk, és még ez is gyorsan eléggé összetett feladat lett.

Az is kiderült, hogy a hálózati architektúránk nem volt eléggé összetett ahhoz, hogy még egy kis szavakból álló tesztkészletet is befogadjon. Bár a bemeneti rétegben rendelkezünk némi kontextusról, nem tettünk lehetővé elégséges kontextust ahhoz, hogy pontos meghatározásokat tegyünk az angol nyelvben lehetséges kontextusok teljes skálájára vonatkozóan.

Egy helyesen működő beszéd szintetizáló hálózat tervezése az angol nyelv teljes egészében a rejtett egységek megtervezésével és a

202
súlyok beállításával

és küszöbértékek kézzel történő meghatározása rendkívül munkaigényes feladat lenne. Szerencsére azonban a mesterséges neurális hálózatok abban is jeleskednek, hogy *megtanulják* a bemenetek és a kimenetek összehangolását.

Bár az előző szakasz hálózata szép példája a mesterséges neurális hálózatok működésének, általában nem szoktunk ilyen módon felépíteni egy érdekes bonyolultságú hálózatot. A rejtett réteg csomópontjainak funkcióját az általunk alkalmazott módon meghatározni, nem felel meg a "rejtett" elnevezésnek. Általában a rejtett réteg csomópontjainak küszöbértékeit és kapcsolati súlyait a hálózat *betanításával* határozzuk meg.

A mesterséges neurális hálózatok számára számos képzési módszer létezik. Az egyik leggyakoribb módszer a *backpropagation* algoritmus alkalmazása a kapcsolati súlyok és küszöbértékek felülvizsgálatára egy generált hibaérték alapján.

A hiba visszaterjedése egy felügyelt képzési módszer, ami azt jelenti, hogy van egy előzetes meghatározásunk arra vonatkozóan, hogy a bemeneteket hogyan kell helyesen leképezni a kimenetekre - pl. a beszédszintetizáló hálózatunkban azt akarjuk, hogy egy adott fonéma kimeneti egysége mindig és csak akkor lépjen működésbe, amikor az adott fonémát az ortográfiai bemeneti kontextus alapján elő kell állítani.

Egy mesterséges neurális hálózat kiképzése beszédszinkronizátorként való működésre a hiba visszaterjedésének segítségével a következőket foglalná magában. Ugyanazokkal a bemeneti poolokkal és kimeneti csomópontokkal kezdenénk, amelyeket az előző szakaszban leírtunk (bár szélesebb bemeneti ablakot - több bemeneti poolt - szeretnénk, hogy több kontextust biztosítsunk). Ezután nagyszámú rejtett csomópontot adnánk hozzá, és minden bemeneti csomópontot minden rejtett csomópontoz és minden rejtett csomópontot minden kimeneti csomópontoz kapcsolnánk (így egy maximálisan összekapcsolt *előreccatolt* architektúrát kapnánk).

A cél az, hogy a hálózat helyesen teljesítsen egy képzési adathalmazon - például a mi szavakból álló tesztkészletünkön. Azzal kezdjük, hogy egyszerűen kis véletlen értékeket rendelünk a kapcsolati súlyokhoz és a küszöbértékekhez, és teszteljük az így kapott teljesítményt. Kezdetben a hálózat nagyon rosszul fog teljesíteni - nem fogja helyesen megfeleltetni a bemeneteket a kimeneteknek - ahogy azt várnánk. Ezután létrehozunk egy hibaértéket, amely jelzi, hogy a hálózat mennyire tért el a helyes leképezéstől. Ezt a hibaértéket ezután visszaterjesztjük a hálózaton keresztül, és a súlyokat és küszöbértékeket a backpropagation algoritmusunknak megfelelően módosítjuk.

A technikai részletekkel itt nem kell foglalkoznunk, mivel a

számítás mérsékelten összetett. A mi céljainkhoz a képzési folyamat fogalmi megértése elegendő. Miután a hálózatot többször is ciklikusan és a hiba visszaterjedése után a hálózat

végül konvergálnak egy olyan állapothoz, amely lehetővé teszi a bemenetek és a kimenetek helyes leképezését a képzési adathalmazon.

Ha a képzési adathalmazunk megfelelően nagy és megfelelően reprezentatív, hogy megfelelően jellemezze a lehetséges leképezések releváns terét, akkor a hálózatnak a képzési halmazon mutatott helyes teljesítménye általánosítható az új bemenetekre. A beszédszintetizáló hálózatunk esetében azt mondhatjuk, hogy a hálózat "megtanulta" az ortográfiai reprezentációk helyes fonémikus átírását.

19.5 MINTAFELISMERÉS

A leírt példánkban, amikor egy mesterséges neurális hálózatot képezünk ki az ortográfia fonémákra való lefordítására, a hálózat megtanulja, hogyan képezze le az ortográfiai összefüggéseket a fonémákra, megtanulva felismerni bizonyos mintákat.

A képzési folyamat során a hálózat mintákat nyer a bemeneti adatokból, és ezeket a számos csomópont közötti kapcsolati súlyok és egységküszöbök formájában kódolja. Más szóval, a képzési adatokban észlelt különböző minták a hálózatban elosztott reprezentációként vannak kódolva.

Bár itt nem mutattam be, a mesterséges neurális hálózatok képesek felismerni (felismerni a tanult minták reprezentációit), még részleges vagy zajos bemenet esetén is. Ez a képesség, hogy az adathalmazokban előforduló mintákat kivonja és kódolja, majd felismeri ezeket a mintákat, amikor azok a későbbi bemenetekben előfordulnak - még akkor is, ha a bemenet nem tökéletes -, teszi a mesterséges neurális hálózatokat alkalmassá a kognitív jelenségek széles körének modellezésére.

Példahálózatunkban ezek a minták bizonyos ortográfiai kontextusok voltak, de a bemenettől függően bármilyen minta lehet. Mintázatok mindenféle környezeti ingerben előfordulnak, és az emberi képesség, hogy érzékeny legyen bizonyos minták előfordulására, alapvetően a kognitív képességek széles skálájához tartozik.

Különösen racionális képességünk függ ettől a képességtől. Intuitív gondolkodásunk nagyon gyakran magában foglalja a múltbeli tapasztalatainkban előforduló, szerkezetileg hasonló helyzetekkel való analógiás összehasonlítást - ez egyfajta mintaillesztés. Emlékezhetünk, hogy még a formális gondolkodási módszerek tudatos követése is megkövetelte, hogy képesek legyünk logikai formákat felismerni, és ezeket a logikai szabályok előzményformáival összevetni.

A mesterséges neurális hálózatok ezen tulajdonságai - a

kontextusra való érzékenységük és a különböző képzési módszerekhez való alkalmazhatóságuk - jó előjelek a mesterséges intelligencia projektekben való sikeres alkalmazásukhoz. Más lényeges előnyökkel is rendelkeznek.

A mesterséges neurális hálózatok könnyen skálázhatók. Például a 19.3. szakaszban bemutatott beszédszintetizáló hálózati részletet tekintve a hálózatot egyszerűen ki lehet bővíteni, hogy más fonémák előállítására vonatkozó meghatározásokat tegyen, vagy hogy a bemeneti réteget szélesebb kontextus figyelembevételével bővítse. Ezt segíti az a tény, hogy a rejtett réteg egységei több feldolgozási funkciót is betölthetnek - például egy bizonyos mintát észlelő egységek gyűjteménye egyszerre gerjeszthet egyes kimeneti egységeket, míg másokat gátolhat.

A mesterséges neurális hálózatok - elvileg - összekapcsolhatók is. Vegyük például az "olvasás" szót. Az ortográfiai kontextus önmagában nem elegendő annak meghatározásához, hogy e szó kimondásakor melyik magánhangzófonémát kell produkálni. Meg kell határozni a *szintaxis* kontextusát is, mivel ez mondja meg, hogy melyik magánhangzót kell produkálnunk. Következésképpen, ha azt akarjuk, hogy a beszédszintetizáló hálózatunk helyesen működjön ebben és hasonló esetekben, akkor együtt kell működnie egy szintaktikai osztályozást végző hálózattal.

Hasonlóképpen, ha egy kész beszédszintetizáló hálózatot szeretnénk, amely képes "természetes" hangzású beszédet produkálni, akkor fonetikai megvalósítási szabályokat kell alkalmaznunk a fonémikus kimenetre. Emellett - ami döntő fontosságú - különböző szemantikai és pragmatikai meghatározásokat is el kell végeznünk, hogy megállapítsuk a mondatok intonációs kontúrjait. Ezek rendkívüli módon diffikult problémák, amelyek megoldhatók számos speciális neurális hálózattal, amelyek párhuzamosan működnek együtt a nyelvi produkció kiszolgálása érdekében.

Végül, a mesterséges neurális hálózatok jellemzően kegyes degradációt mutatnak, hasonlóan az emberi agyakhoz. Egyetlen elem - egy regiszter vagy egy kódsor - eltávolítása egy regisztergépből általában elég ahhoz, hogy az teljesen tönkremenjen. A mesterséges neurális hálózatok ezzel szemben sokkal robusztusabbak a sérülésekkel szemben. Kis számú elem eltávolítása kevés vagy semmilyen hatással nem járhat. A káros effektusok, ha további sérülések következtében keletkeznek, a sérült hálózat újratanításával javíthatók úgy, hogy az visszanyeri funkcióit - ahogyan a szélütéses betegek is újratanulják a kognitív funkciókat.

19.6 KÉT PARADIGMA?

Bár a mesterséges intelligencia szimbolikus és konnektionista megközelítését alapvetően különböző - és ebből következően összeegyeztethetetlen - *paradigmáknak* neveztem, könnyen lehet,

hogy ezek az információfeldolgozással kapcsolatos nézetek csupán a leírás különböző szintjein mozognak.

A konnekcionista paradigmát gyakran nevezik *szubszimbolikus* paradigmának, ami arra utal, hogy a szimbólumrendszereknél alacsonyabb szintű leírással foglalkozik. Alternatív megoldásként az is lehet, hogy az emberi megismerésben az alacsony szintű szimbolikus feldolgozás bizonyos fajtái a magasabb szintű konnekcionista feldolgozást szolgálják.

Első látásra úgy tűnik, hogy a mesterséges neurális hálózatok működése (legalábbis ahogyan itt leírtuk) teljesen effektív. Ezért a Church-Turing tézis szerint ezek a gépek regisztrálhatóak. Természetesen az átviteli függvények és az aktiválási függvények algoritmikusak, és úgy tűnik, hogy a párhuzamos feldolgozást lépéenkénti soros feldolgozással közelíthetjük meg, így talán a konnekcionizmus egyszerűen szimbolikus feldolgozásra redukálódik.

Másrészt láttuk, hogyan lehet logikai kapukat építeni mesterséges neurális hálózatokkal. Az általunk ismert számítógépek alapvetően logikai kapukból épülnek fel, így talán a szimbólumrendszerek egyszerűen visszavezethetők a konnekcionista feldolgozásra.

A gyakorlatban a szimbolikus modellek és a mesterséges neurális hálózatok modelljei egyáltalán nem összeegyeztethetetlenek, mivel a mesterséges neurális hálózatokat szimulálják a szimbólumrendszerek architektúráján. A neurális biomérnöki tudományok születőben lévő és fejlődő területén a közelmúltban elért eredmények azonban kiveszik a szimbólumrendszereket az egyenletből. A valódi neuronokból felépített biológiai neurális hálózatokról kimutatták, hogy a mesterséges neurális hálózati modellek számos jellemzőjét mutatják - beleértve azt a képességet, hogy bizonyos komplex funkciók végrehajtására, például egy repülőgép-szimulátor szakszerű működtetésére is képesek.

19.7 EZ CSAK EGY MODELL

Ebben a fejezetben a mesterséges neurális hálózatokkal kapcsolatos bevezetés valóban nagyon alapszintű volt. A felesleges matematikai bonyolultság elkerülése érdekében csak a legegyszerűbb típusú hálózatokat és függvényeket vettük figyelembe. Nyomatékosan ajánlom az érdeklődő olvasónak, hogy folytassa vizsgálódásait a további olvasmányokra vonatkozó javaslatok segítségével. A mesterséges neurális hálózatok megfelelő bevezetéséhez külön tankönyvre van szükség.

A mesterséges neurális hálózatok modelljei minden kifinomultságuk és összetettségük ellenére is a modellezni kívánt biológiai idegi tevékenység durva leegyszerűsítései maradnak. Különösen nem veszik figyelembe a neurotranszmitterek globális és analóg effektusait, és ennek mélyreható következményei

vannak számos mentális jelenség, köztük (döntően) a figyelem és az érzelmek modellezésének lehetőségére.

Ahogy azonban egyre többet tudunk meg az agyról, talán még kifinomultabb modelleket tudunk kidolgozni, amelyek megvalósítják az általunk feltárt neurobiológiai elveket. Különösen érdekes lesz látni, hogy az idegrendszeri biomérnöki tudományok fejlődése a következő évtizedben nyújt-e empirikus táplálékot a számítási neurális modellezéshez.

ELMÉK ÉS SZÁMÍTÓGÉPEK

Mostanra már sokat tanultunk az elméről, miután áttekintettük a rendelkezésre álló filozófiai elméleteket, és megvizsgáltuk az egyes elméletek előnyeit és hátrányait.

Sokat tanultunk a számítógépekről is, hiszen kidolgoztunk egy pontos technikai leírást arról, hogy pontosan mi is az a számítógép, és gyakoroltuk a számítógépes programozás alapjait.

Láttuk, hogyan alkalmazhatunk szimbólumrendszereket a megismerésben szerepet játszó számos funkció megvalósítására - különösen a racionális és nyelvi képességek tekintetében.

Útközben megtanultunk néhány alapvető funkcionális neuroanatómiát, egy kis formális logikát, egy kis nyelvészetet, és a modern kognitív pszichológia rövid érintése mellett megismerkedtünk az empirikus pszichológia korai történetével is.

Végül megnéztünk néhány egyszerű mesterséges neurális hálózatot, és láttuk, hogyan alkalmazhatunk ilyen konnekcionista hálózatokat a kognitív jelenségek modellezésében - ismét különös tekintettel a racionális és nyelvi képességekre.

Mindez az erős mesterséges intelligencia projektjének tarthatóságát vizsgáló interdiszciplináris vizsgálat szolgálatában állt.

Ebben az utolsó fejezetben csak röviden szeretnék kitérni a mesterséges intelligenciával kapcsolatos néhány filozófiailag "nehéz" problémára - nevezetesen a tudattal, a személyes identitással és az érzelmekkel kapcsolatos problémákra.

20.1 TUDATOSSÁG

Bár helyenként segítettem magamnak egy intuitív megkülönböztetéshez a tudatosan észlelt mentális folyamatok és a tudat szintje alatt zajló folyamatok között, a tudatosságról önmagában nem sokat beszéltem.

Az emberi mentalitással kapcsolatos kiküszöbölhetetlen - de talán nem is megmászható - tény, hogy van *tudatunk*. A "tudatos" szót azonban sokféleképpen használják.

Néha arra használják, hogy bizonyos események vagy folyamatok tudatosságára utaljon, *amelyeknek tudatában* vagyunk. Néha az önmagunkkal kapcsolatos tudatosságunkra, valamint az önmagunk és a világ többi része közötti különbségtételre - az *öntudatunkra* - utal. Néha pusztán az ébrenléti és az alvási - vagy *tudattalan* - állapotok megkülönböztetésére használják.

Bizonyos vallásokban a *tudatosság* azt jelenti, hogy valaki ensouled. A pszichoanalitikus elméletben általában megkülönböztetik a *tudatos* elmét a *tudatalatti* elmétől, és ezeket jellemzően mindenféle feszültségben lévőknek tartják, amelyet csak a pszichoanalízis képes feloldani.

Filozófiai szempontból a tudatosság magában foglalja a szubjektív tapasztalás képességét és az ehhez kapcsolódó, kiváltságos, első személyű tapasztalati tulajdonságok - *a kvaliák* - meglétét. Szintén erősen összefügg az intencionális tartalommal bíró reprezentációs állapotok kialakításának képességével.

Kevésbé egyértelmű, hogy egyetlen átfogó "tudatossági problémáról" van-e szó, vagy több releváns problémáról - talán "könnyebb" és "nehezebb" problémákról -, bár David Chalmers sokat tett az érzékek tisztázása és a filozófiai kérdések szétválasztása érdekében.

A tudatosság jelenleg az elmefilozófia legforróbb témája, és a jelenség vizsgálatára külön kutatóközpontok alakultak. Ezek a kutatóközpontok olyan interdiszciplináris elemzésekkel foglalkoznak, mint amilyeneket ebben a kötetben végeztünk, és nagy hangsúlyt fektetnek arra, hogy pontosan meghatározzák, melyek a releváns filozófiai kérdések, és hogyan lehet rájuk válaszolni.

A filozófiai haladó olvasóknak érdemes követniük a további olvasmányokra vonatkozó javaslatokat, hogy jobban megértsék a filozófia e kihívást jelentő, izgalmas és fejlődő területét.

20.2 SZEMÉLYES IDENTITÁS

Egy adott napon egyértelműen több szempontból is differ vagyok, mint előző nap, mivel számos differ tulajdonságom lesz. Különleges tér-időbeli elhelyezkedésem lesz, megváltozott vagy kibővített hiedelmeim lehetnek, extra emlékeim lesznek, testem apró darabjai - bőr, haj, köröm és hasonlók - elvesznek, és újak nőnek, és így tovább.

E számos különböző tulajdonságom ellenére azonban napról napra és évről évre mindig *ugyanaz* az *ember* vagyok - van egy

204
egyedülálló

személyes identitás, amely tér-időbeli, pszichológiai és anyagi tulajdonságaim számos változásán keresztül fennmarad. Bár minőségileg napról napra különbözöm, számszerűen mégis azonos vagyok - azaz egy és ugyanaz a személy.

Nem nehéz problematizálni a tartós személyes identitás fogalmát.

Függetlenül attól, hogy milyen kritériumokat részesítünk előnyben a személyes identitás időbeli fennmaradásával kapcsolatban, úgy tűnik, hogy problémás eseteket tudunk felhozni. Gyakori, hogy a pszichológiai folytonosságot részesítjük előnyben, mint a személyes identitás fennmaradásának kritériumát. A pszichológiai folytonosság azonban nem tűnik a személyes identitás fennmaradásának *szükséges* feltételének, hiszen elképzelhető, hogy teljes amnéziám van, úgy, hogy pszichológiai állapotaim radikálisan megszakadnak a múltbeli pszichológiai állapotokkal. Intuitív módon azonban még mindig *ugyanaz* a *személy* lennék - csak elvesztettem az emlékezetemet. Van egy-két válasz

itt elérhető, de ezt az olvasóra bízom.

A pszichológiai folyamatosság kritériuma sem tűnik *megfelelőnek* a személyes identitás fennmaradásához. Tegyük fel, hogy beelépek egy anyagteleportáló eszközbe, és valamilyen baleset folytán a másik végén kétszer is összeraknak. Mindkét lény, aki kilép az anyagtransz- portálóból, pszichológiailag folytonos velem - legalábbis abban a pillanatban, amikor kilépnek -, tehát, ha a pszichológiai folytonosság a személyes identitás fennmaradásának megfelelő kritériuma, akkor mindkettőnek ugyanannak a személynek kell lennie, mint én. A válaszokat ismét az olvasóra bízom.

Most tegyük fel, hogy a komputacionalizmus igaz. Ez azt jelenti, hogy elvileg pontosan le tudnám másolni az elmédet az agyadon kívüli számítási hardverrel. Tegyük fel, hogy van egy olyan számítási eszközöm, amely kellően erős ahhoz, hogy ugyanolyan jól működtesse az [ÉN]-edet, mint az agyadat. Tegyük fel továbbá, hogy egy éjjel, miközben Ön alszik, egy új, csodás szkennelvel és néhány innovatív sebészeti technikával letapogatom az Ön [ELMÉLETÉT], lemásolom azt a számítási eszközömben, majd kicserélem az agyát az új számítási eszközzel anélkül, hogy Ön valaha is észrevenné.

Mit ság az intuíciója ebben az esetben? Még mindig ugyanaz a személy vagy? Ha úgy gondolja, hogy nem, akkor módosítsa a példát úgy, hogy az első éjszakán csak az egyik neuronját cserélem ki egy mesterséges neuronra. A következő éjszakán pedig tízet cserélek ki. A következő éjszakán kicserélek száz, majd ezer, majd egymillió, és így tovább - mindezt anélkül, hogy te valaha is észrevennéd. Hacsak nem vagy hajlandó jelezni, hogy a kicserélt neuronok hány darabja jelenti a személyes identitásod változását,

úgy tűnik, hogy a folyamat végén ugyanannak a személynek kell lenned. Ha úgy gondolod, hogy ugyanaz a személy vagy az agyadat helyettesítő alternatív kommunikációs eszközzel, akkor módosítsd a példát így

hogy én csupán letapogatom az [ELMÉT], majd az azt megvalósító számítógésséget egy android testbe helyezem, és úgy hagylak, ahogy vagy. Mit mondanak most az intuícioid? Milyen kötelezettségeink vannak, ha vannak egyáltalán, az android testtel és a te [ÉNELEDdel] szemben?

A további olvasmányajavaslatok közé felvettem néhány olyan cikket, amelyek a személyes identitást problematizálják, ugyanakkor a bevezető olvasó számára is hozzáférhetőek és szórakoztatóak maradnak.

20.3 EMOTIONS

Általánosságban úgy tartják, hogy az érzelmi állapotok és reakciók képességének hiánya az elme meglétének egyik feltétele, abban az értelemben, hogy van elménk. Az érzelmi viselkedés hiánya bizonyos pszichopatológiák egyik jellegzetes tünete, és bár az ilyen embereket elmével rendelkezőnek tartjuk, úgy véljük, hogy az ő elméjük minőségileg jelentősen különbözik a miénktől.

A sci-fiben szinte mindig az érzelmi elkötelezettséget használják arra, hogy elmosódjon a határ az emberek és a mesterséges intelligencia között. Gondoljunk csak a *Bladerunner*, a *Terminator II* és az *I, Robot* című filmek befejezésére. Mindegyik esetben érzelmi ellenszenvvel viseltetünk a mesterséges intelligenciával rendelkező főszereplő iránt azért, hogy elhisszük, hogy az képes érzelmeket táplálni.

Az érzelmek a mentalitás egyik legkevésbé jól megértett aspektusa. Tudjuk, hogy bizonyos érzelmek bizonyos neurotranszmitter-kombinációkkal korrelálnak és ~~mindkét~~ e folyamatok megértése a legjobb esetben is csak vázlatos. Azt is tudjuk, hogy az agy bizonyos lokalizált területeinek károsodása jellegzetes érzelmi hiányosságokat eredményezhet.

Az egyik különösen érdekes dolog, amit az érzelmekről tudunk, hogy az érzelmi elkötelezettség erősen kötődik az epizodikus emlékezethez. Nyilvánvalóan az a helyzet, hogy sokkal nagyobb valószínűséggel emlékszünk vissza olyan eseményekre, amelyek erős érzelmi reakciót váltottak ki belőlünk. Továbbá tudjuk, hogy az agy limbikus rendszere az érzelmekben és az emlékezeten egyaránt szerepet játszik.

Intuitívan világos, különösen az általam említett sci-fi példák tükrében, hogy sokkal valószínűbb, hogy azt hisszük, hogy egy mesterséges intelligencia, amely az emberi érzelmi reakciók teljes skálájával rendelkezik, ugyanolyan értelemben elmével rendelkezik, mint mi. Itt azonban ismét felmerül a probléma, hogy egy másik elme minőségi aspektusai nem hozzáférhetőek a tapasztalathoz.

Ha egy mesterséges intelligencia az emberi érzelmi reakciók

szokásos skáláját mutatná, de ezek csak külsődleges megnyilvánulások lennének, melyek

a mesterséges intelligencia nem *érezne* semminek, akkor is azt a robusztus fogalmat tulajdonítanánk neki, hogy van elméje? Ha nem, akkor miért tulajdonítunk elmét más embereknek, amikor az érzelmi állapotukról csak a megfigyelhető viselkedésükből tudunk következtetni?

Mint mindig, most sem egyértelmű, hogy mit kellene mondanunk a kvaliákról, és ezt az olvasóra bízom, hogy mérlegelje.

20.4 SZÁMÍTÓGÉPEK ELMÉVEL

Most, hogy a könyv végére értünk, ideje elgondolkodni azon, hogy milyen megállapításokat tehetünk a mesterséges intelligencia lehetőségével kapcsolatban.

Nem láttunk itt semmi olyat, ami arra engedne következtetni, hogy az erős mesterséges intelligencia *lehetetlen*, bár láttunk néhány kiindulópontot ilyen érvek felhozatalához. Prima facie, engedve a további filozófiai kutatások lehetséges meghatározásainak, úgy tűnik, hogy lehetséges lehet olyan számítógépet tervezni, amelynek van elméje abban az értelemben, ahogy nekünk is van elménk.

Láttuk azonban, hogy a megismerés jelenlegi legjobb számítógépes modelljei még mindig szánalmasan elégtelenek, de reménykedünk abban, hogy az idegtudományok fejlődése technikai megértést biztosíthat számunkra a megismerést szolgáló biológiai folyamatokról, ami gazdagabb fogalmi megértéshez és végül sikeres, erős mesterséges intelligencia projektekhez vezethet.

Sikerült a mesterséges intelligencia fejlődésének néhány feltételes szükségességnek tűnő feltételt szabnunk. A 17. fejezetben amellet érveltünk, hogy a megtestesült tapasztalat szükséges feltétele a szemantika kialakulásának, amely viszont szükséges ahhoz, hogy elmével rendelkezünk.

Következésképpen, ha mesterséges intelligenciát akarunk kifejleszteni, akkor annak elsősorban a külső világhoz kell kapcsolódnia a megfelelő módon. Más szóval, olyan érzékszervekkel kell rendelkeznie, amelyek közvetítik a közte és a külvilág közötti kapcsolatokat. Továbbá, embrionális mesterséges intelligenciánknak képesnek kell lennie arra, hogy tapasztalatokat gyűjtsön, amelyek révén mentális reprezentációkkal ruházzuk fel.

Tekintettel az elme jelenlegi fogalmi megértésére és az azt létrehozó agy számítógépes "wetware"-jének technikai megértésére, messze a legegyszerűbb módja annak, hogy létrehozunk valamit, ami képes a megtestesült tapasztalatszerzésre, és ami *ceteris paribus* garantáltan kifejleszti az elmét ugyanabban az értelemben, ahogy nekünk is van elménk, még mindig a régimódi biológiai módszer - egy emberi lény létrehozása és felnevelése.

I. FÜGGELÉK: TOVÁBBI OLVASMÁNYJAVASLATOK

2. FEJEZET

Campbell, K. *Test és elme*. London: Churchland, P. *Matter and Consciousness: Matter and Consciousness (Anyag és tudat)*.

Cambridge, MA: MIT .

Press, 1988, 2. fejezet.

Descartes, *Elmélkedések az első filozófiáról*, ford. J. Cottingham.

Cambridge: Cambridge University Press, 1986, pp. 50-6.

3. FEJEZET

Campbell, K. *Test és elme*. London: Gardner, H. *Az elme új tudománya*. New York: Basic Books, 1985,

pp. 98-114.

Ryle, G. *Az elme fogalma*. Harmondsworth: Penguin, 1973, pp. 13-25.

Schultz, D. *A modern pszichológia története*. New York: Academic Press, 1975, 3., 4., 5., 10., 11. fejezet.

4. FEJEZET

Barr, M. *Az emberi idegrendszer: An Anatomic Viewpoint*, 3rd edn. Hagerstown, MD: Harper & Row, 1974.

Diamond, M. C. et al. *The Human Brain Coloring Book*. New York: Harper Collins, 1985.

Gregory, R. (szerk.): *The Oxford Companion to the Mind*, Oxford: Oxford University Press, 1987, pp. 511-60.

5. FEJEZET

Armstrong, D. *Az elme materialista elmélete*. London: London: Routledge & Kegan Paul, 1968.

- Jackson, F. "Epiphenomenal Qualia", *Philosophical Quarterly*, 32 (1982), pp. 127-36.
- Nagel, T. "Milyen érzés denevérenek lenni?", *Philosophical Review*, 83 (1974), pp. 435-50.
- Place, U. T. "Is Consciousness a Brain Process?", *British Journal of Psychology*, 47 (1956), pp. 44-50.
- Smart, J. J. C. "Sensations and Brain Processes", *Philosophical Review*, 68 (1959), pp. 141-56.

6. FEJEZET

- Block, N. "Troubles with Functionalism", reprinted in Block, N. (szerk.) *Readings in Philosophy of Psychology*, 1. kötet. Cambridge, MA: Harvard University Press, 1980.
- Guttenplan, (szerk.) *A Companion to the Philosophy of Mind*. Oxford: Blackwell, 1994, pp. 317-32.

7-9. FEJEZET

- Church, A. "Az elemi számelmélet megoldhatatlan problémája", *American Journal of Mathematics*, 58 (1936), pp. 345-63.
- Hofstadter, D. *Gödel, Escher, Bach: Egy örök aranyfonat*, 20. évfordulós kiadás. London: Penguin, 2000, pp. 33-41.
- Jeffrey, R. *Formális logika: A logikai logika: Terjedelme és határai*, 3. kiadás. New York: McGraw-Hill, 1991, pp. 100-2.
- Turing, A. M. "On Computable Numbers, with an Application to the Entscheidungsproblem", *Proc. London Math. Soc.*, 42 (1937), pp. 230-65.

10. FEJEZET

- Scheutz, M. (szerk.) *Computationalism: New Directions*, Cambridge, MA: MIT Press, 2002, 1. fejezet.
- Turing, A. M. "Computing Machinery and Intelligence", *Mind*, 59 (1950), pp. 433-60.

11-13. FEJEZET

- Copeland, J. *Mesterséges intelligencia: A Philosophical Introduction*. Oxford: Blackwell, 1993, 4. fejezet.
- Haugeland, J. *Mesterséges intelligencia: The Very Idea*. Cambridge, MA: MIT Press, 1989, 2. fejezet.

- Newell, A. és Simon, H. "A számítástechnika mint empirikus kutatás: *Communications of the ACM*, 19 (1976), 113-26. o., "Symbols and Search", *Communications of the ACM*, 19 (1976).
Pinker, S. *Hogyan működik az elme*. London: Allen Lane, 1998, 2. fejezet.

14. FEJEZET

- Chomsky, N. *Aspects of the Theory of Syntax*. Cambridge, MA: MIT Press, 1965.
Haegeman, L. *Kormányzat és kötélmélet*. Oxford: Blackwell, 1991.

15. FEJEZET

- Cohen, L. J. "Can Human Irrationality be Experimentally Demonstrated?", *Behavioural and Brain Sciences*, 4 (1981), pp. 317-30.
Gardner, H. *Az elme új tudománya*. New York: Basic Books, 1985, 13. fejezet.
Johnson-Laird, P. N. *Mental Models: Towards a Cognitive Science of Language, Inference and Consciousness*. Cambridge, MA: Harvard University Press, 1983.
Kahneman, D., Slovic, P. és Tversky, A. (szerk.) *Judgment under Uncertainty: Heuristics and Biases*. Cambridge: Cambridge University Press, 1982.
Wason, P. C. "Natural and Contrived Experience in a Reasoning Problem", *Quarterly Journal of Experimental Psychology*, 23 (1971), pp. 63-71.

16. FEJEZET

- Crystal, D. *The Cambridge Encyclopedia of the English Language*. Cambridge: Cambridge University Press, 1995, 17. fejezet.
Fromkin, V., Rodman, R. et al. *An Introduction to Language*, 2. kiadás. Sydney: Holt, Rinehart & Winston, 1990, 2. és 3. fejezet.

17. FEJEZET

- Searle, J. "Minds, Brains and Programs", *Behavioural and Brain Sciences*, 3 (1980), pp. 417-57.

18. FEJEZET

- Fodor, J. *Fogalmak: Where Cognitive Science Went Wrong*. Oxford: Clarendon, 1998.
- Putnam, H. *Értelem, igazság és történelem*. Cambridge: Cambridge University Press, 1981, 1. fejezet.
- Smith, E. és Medin, D. *Kategóriák és fogalmak*. Cambridge, MA: MIT Press, 1981.

19. FEJEZET

- Dennis, S. és McAuley, D. *A megismerés konnekcionista modelljei*. Online szöveg, a cikk írásakor elérhető a következő címen: <http://lsa.colorado.edu/~simon/cmc/index.html>.
- Rumelhart, D. és McClelland, J. et al. *Parallel Distributed Processing: Explorations in the Microstructure of Cognition. Vol. 1: Foundations*. Cambridge, MA: MIT Press, 1986, 1., 2., 3., 5., 8. fejezet.

20. FEJEZET

- Chalmers, D. *A tudatos elme*. New York: Oxford University Press, 1996.
- Dennett, D. *Agyviharok: Philosophical Essays on Mind and Psychology*. Cambridge, MA: MIT Press, 1981.
- Hofstadter, D. és Dennett, D. (szerk.) *The Mind's I*. New York: A tudat énje: Basic Books, 1981, 26. fejezet.

II. FÜGGELÉK: A KIFEJEZÉSEK SZÓJEGYZÉKE

a fortiorieven erősebben, ugyanazzal az érveléssel. "Brunswick Melbourne-ben van, tehát Victoriában van; *a fortiori*, Ausztráliában van." **Aktiválási függvény** Egy *mesterséges neurális hálózat* minden egyes feldolgozó csomópontja által végrehajtott két függvény egyike, amely meghatározza, hogy az aktiváltsági szint alapján a csomópont tüzelni kellene.

adicity a *függvény* által felvett bemenetek száma.

affkülönböző bejövő (jelvezetés).

algoritmus egy másik elnevezés az *effektív* eljárásra - olyan szisztematikus módszer, amelynek végrehajtásához nincs szükség megértésre, és amely véges idő alatt elvégezhető.

allofóna egy *fonéma* fonetikai változata, amelynek kontextusfüggő produkcióját *fonetikai megvalósítási szabályok* határozzák meg. **amphibolya** bizonyos *szintaktikai* szerkezetek olyan tulajdonsága, hogy többféle *szemantikai* értelmezés elismerése. "Láttam a távcsöves embert a dombon.

anomális monizmus az a nézet, hogy bizonyos fizikai állapotok irreálisan nem fizikai tulajdonságokkal rendelkeznek.

antecedens a *feltételes mód* bal oldala.

afázia nyelvi deficit. Gyakori formái a Broca-féle afázia, a Wernicke-féle afázia és a vezetési afázia.

a mesterséges intelligencia, az a vitatott nézet, hogy lehetséges olyan tárgyakat fejleszteni, amelyeknek olyan értelemben van elméjük, mint nekünk van elménk.

mesterséges intelligencia, gyenge az a nem vitatott nézet, hogy lehetséges olyan műtárgyakat létrehozni, amelyek képesek bizonyos funkciókat végrehajtani, amelyeket (gyengén) az intelligencia konstitutív elemeinek tartanak. Gyakran használják fehérráruk eladására.

mesterséges neurális hálózat egyszerű feldolgozóegységek közötti súlyozott kapcsolatok struktúrája - amelyek mindegyike egy *átviteli függvényt* és egy *aktiváló függvényt* valósít meg - a

212

terjedés a

211

aktiválás, amely párhuzamos elosztott feldolgozást jelent. A mesterséges neurális hálózatokat az "agy stílusú" információfeldolgozásra modellezzik.

behaviorizmus, filozófiai (analitikus) az elme szemantikailag reduktív és ontológiailag eliminatív elmélete, amely szerint a mentális állapotokra vonatkozó kifejezéseink valójában nem választanak ki semmit, hanem egyszerűen csak hasznos szűkszavú kifejezések a viselkedési diszpozíciók komplex halmazaira való utaláshoz.

behaviorizmus, pszichológiai módszertani nézet a pszichológiai vizsgálat menetére vonatkozóan, amely szerint a pszichológiának kizárólag a megfigyelhető viselkedéssel kell foglalkoznia. Befolyásolta a *pozitivizmus*.

kauzális elmélet (az elme) az a nézet, amely szerint a mentális állapotokat kizárólag az inger és a viselkedés közötti közvetítő kapcsolatok tekintetében betöltött kauzális szerepük alapján jellemzik.

ceteris paribus minden más dolog egyenlő.

A kínai szoba egy gondolatkísérlet, amely azt igyekszik megmutatni, hogy a külső világtól elszigetelten semmilyen *szintaktikai* manipuláció nem alkalmas a *szemantika* létrehozására.

kognitív architektúra a kognitív ágens információfeldolgozó rendszereinek felépítése és jellege - a kognitív ágens információfeldolgozó hardverének szervezési és végrehajtási jellemzői.

számítás egy *regisztergép* műveletsorozata.

komputacionalizmus az elme filozófiai elmélete, amely szerint az elme olyan, mint az agy hardverén, vagy *wetware*-jén futó szoftver - egyfajta *funkcionalizmus*, amely lehetővé teszi a számítógépes *mesterséges intelligencia* lehetőségét.

számítógép egy *univerzális gép* fizikai megvalósítása - pontosabban megközelítése -.

feltételes egy "Ha akkor . . ." formájú állítás.

konjunkció egy logikai művelet, amelyet a természetes nyelvi "és" kifejezéssel ábrázolnak.

következmény a *feltételes feltétel* jobb oldala.

ellenpélda egy olyan értelmezés, amely megmutatja, hogy egy következtetési forma nem *érvényes* azáltal, hogy megmutatja, hogy lehetséges, hogy az összes premissza igaz, míg a konklúzió hamis.

eldönthető bizonyos *formális* rendszerek olyan tulajdonsága, hogy létezik egy *effektív* eljárás annak meghatározására, hogy a rendszer bármely adott *állapota generált-e* vagy sem.

dedukció olyan következtetés, amely valamilyen logikai elvre, például

a *modus ponens*re hivatkozik.

levezetés egy olyan bizonyítás, hogy valamilyen *formális rendszer* egy adott *állapota generálódik*, amely a rendszer szabályainak előírt alkalmazási sorrendjét mutatja.

determinisztikus egyes *formális rendszerek* olyan tulajdonsága, hogy legfeljebb egy *szabály* alkalmazható bármely adott *állapotra*, és csak egyféleképpen.

diphthonga olyan *hangzó fonéma*, amelynek előállításához a nyelvnek az egyik kardinális magánhangzó pozícióból a másikba történő mozgatása szükséges. **megkülönböztető kontrasztok**
hangzó megkülönböztető kontrasztot mutat - és a két
hangzó

ezáltal különböző *fonémák* - *iff* az egyik *telefon* egy másikkal való helyettesítése jelentésváltozást eredményez.

dualizmus, karteziánus az elme filozófiai elmélete - más néven interakcionista dualizmus, amely szerint az elme és a test különböző szubsztanciákból áll, amelyek kölcsönhatásban állnak egymással.

dualizmus, szubsztancia olyan *ontológiai* nézet, amely szerint az egység két különböző szubsztanciából - fizikai és nem fizikai, vagy anyagi és nem anyagi - áll.

effektivitás bizonyos eljárások - más néven *algoritmusok* - olyan tulajdonsága, hogy szisztematikusan, véges idő alatt, megértés nélkül végrehajthatók. Az effektivitás korlátozza a *formális rendszerek állapotait* és *szabályait*.

efferent kimenő (jelvezetés).

epifenomenalizmus az elme filozófiai elmélete, amely szerint a mentális állapotok nem fizikai, de kauzálisan inaktívak - bizonyos fizikai folyamatok pusztán epifenoménei.

szakértői rendszer egy különleges *formális rendszer*, amelyet olyan *következtetések generálására* használnak, amelyek egy adott szakértő érvelési folyamatait reprezentálják.

exponenciálás az a matematikai művelet, amelynek során egy számot egy másik szám hatványára emelnek, és amelynek példái a négyzetelés és a köbösítés.

formális rendszer *effektíven* megkülönböztethető *állapotok* gyűjteménye és olyan *szabályok* gyűjteménye, amelyek *effektíven* működnek az állapotokon, hogy más állapotokat *hozzanak létre*. A társasjátékok, mint például a sakk, paradigmás *formális rendszerek*.

függvény egy matematikai összefüggés néhány fix számú bemenet és néhány egyedi kimenet között.

funkcionalizmus az elme filozófiai elmélete, amely szerint a mentális állapotok meghatározó jellemzői a bemenetek, kimenetek és más mentális állapotok közötti kapcsolatok közvetítésében betöltött funkcióik.

generálás az a folyamat, amelynek során egy *formális rendszer* egy *állapotára* egy *szabályt* alkalmazunk, hogy egy másik állapotot kapjunk.

generációs fa egy struktúra, amely az alkalmazások sorozatát mutatja a

szabályok egy formális rendszer állapotaihoz.

generatív nyelvtan egy *formális rendszer*, amely egy nyelv karakterlánccainak nyelvtani jellegéről dönt.

Gödel a prímszámok exponenciálásával kódolt egy módszert, amely megkönnyíti a *formális rendszerek* elemeire való hivatkozást a rendszeren belülről.

heurisztika (funkció) informálisan, a megoldáshoz vezető módszer. Formálisan a *heurisztikus függvények egy formális rendszer állapotaihoz olyan értéket rendelnek*, amely az adott állapotnak valamilyen célállapothoz való közelségét jelzi.

heurisztikus keresés olyan keresési módszer, amely egy *heurisztikus függvény* által informált.

homonímia olyan *homofon*, amelynek több szemantikai értelmezése is osztozik egy ortográfiai ábrázoláson - "bank" és "bank".

homofon olyan fonémák sorozata, amely egynél több *szemantikai* értelmezést tesz lehetővé - "tenyésztett" és "kenyér".

identitáselmélet (az elme) az ausztrál/reduktív/központi állam materializmusának másik neve.

iff ha és csak ha.

az indukció az empirikus tudományokra jellemző következtetési forma, amely szerint a premisszák igazsága alátámasztja a következtetést, de nem garantálja azt.

szándékosság a mentális állapotok azon tulajdonságának filozófiai szakkifejezése, hogy valamit képviselnek vagy valamiről szólnak.

többek között.

ipso facto éppen ezzel a ténnyel.

izomorfizmus Az egyik *formális rendszerből* egy másikba történő *izomorfizmus* a rendszer elemeinek olyan struktúrakarbantartó, egységes cseréje, amely során minden formális tulajdonság megmarad.

ismeretelméleti érv olyan, gondolati kísérletből levont érv, amely azt igyekszik megmutatni, hogy a mentalitás teljes fizikai magyarázata nem jelent teljes magyarázatot.

Loebner-díj a *Turing-teszt* bevezetett és évente megrendezett végrehajtása.

logika a *logikát* vizsgáló kutatási program **logikája**.

logikák *formális rendszerek*, amelyek a következményi relációkat kódolják - mi "következik" miből, mint logikai forma kérdése.

materializmus, reduktív/ausztrál/középpontú állapot különböző elnevezései annak a filozófiai elméletnek, amely szerint a mentális állapotok típusai azonosak az idegi állapotok típusaival.

minimax egy eljárás a nyerő stratégiák meghatározására egy kétjátékos játékot reprezentáló *formális rendszerrel* összefüggésben.

modus ponens az a logikai elv, amely szerint egy *feltételes feltétel előfeltételének* igazságából *következtethetünk a feltételes feltétel következményének* igazságára.

modus tollens az a logikai elv, amely szerint egy *feltételes feltétel következményének* hamisságából *következtethetünk annak előzményének* hamisságára.

monophthong olyan *hangzó fonéma*, amelynek előállításához a nyelvnek a kardinális magánhangzók egyikénél kell tartania a nyelvet.

a mentális állapotok **többszörös megvalósíthatósága**, azaz az a tény, hogy több fizikai megvalósításuk is lehet, akár alanyok között, akár egy alanyon belül, időben.

neurális plaszticitás az emberi agy azon tulajdonsága, hogy az agy bizonyos részei képesek átvenni az agy sérült részeinek megszokott funkcióját (funkcióit) - különösen a fiatalabb agyaknál jellemző.

neuron az idegrendszer egyes sejtjei.

normatív tágabb értelemben olyan elv vagy elvek összessége, amely azt hivatott megmondani, hogy mit kellene tennünk.

obstruens olyan *fonéma*, amelynek előállításához a levegőnek az artikulációs készüléken való áthaladása részben vagy teljesen akadályozva van.

occasionalizmus egyfajta elme-test *dualizmus*, amely szerint az elme és a test nem lép kölcsönhatásba egymással, hanem Isten lép közbe időről időre, hogy a dolgok úgy tűnjenek, mintha így lenne.

Ockham borotvája az elméletalkotás módszertani korlátja, amely azt állítja, hogy nem szabad több entitást feltételezni, mint amennyi szigorúan szükséges a magyarázathoz.

ontológia a metafizika magja, amely azzal foglalkozik, hogy mi van alapvetően - a *dualizmus* ontológiája az, amely szerint két különböző szubsztancia létezik.

operáns kondicionálás a Skinner által leírt kondicionálás, amely abban különbözik a *pavlovi* kondicionálástól, hogy a viselkedést követő inger is kondicionáló hatással bír.

a **párhuzamosság** egyfajta elme-test *dualizmus*, amely szerint az elme és a test nem lép kölcsönhatásba egymással, de Isten a dolgokat eredetileg - előre elrendelt harmóniában - úgy állította be, hogy úgy tűnjön, mintha így lenne.

Pavlovi kondicionálás az a folyamat, amelynek során bizonyos ingerek megbízhatóan kiváltanak bizonyos viselkedést.

önmagában véve.

petitio principii a kérdés felvetésének tévedése, amikor az ember

feltételezi, hogy a kérdéses állítást elismeri.

fonémák idealizált kategóriák - amelyekhez a *telefonok* asszimilálódnak -, amelyek egy nyelvben az atomi értelemmel bíró beszédhangok állományát képviselik.

telefonok fonetikusán megvalósított beszédhangok.

fonetikai realizációs szabályok azok a szabályok, amelyek szerint a *fonémák allofón* változatai kontextusfüggő komplementer eloszlásban keletkeznek.

mondatszerkezeti nyelvtan egy olyan *determinisztikus formális rendszer*, amelyben minden *szabály* kontextusmentes és pontosan egy szimbólumot tartalmaz a bemeneti oldalon.

phrase structure tree egy mondatszerkezeti nyelvtan *generációs* fája.

pozitívizmus a tizenkilencedik század végén és a huszadik század elején befolyásos tévtan, amely szerint a "valódi" vagy "pozitív" tudománynak kizárólag a megfigyelhető entitásokkal kell foglalkoznia.

predikátum a nyelvészetben, egy ige az összes kiegészítésével együtt. E kötet céljaira a *predikátumot úgy is* felfoghatjuk, mint ami egy tulajdonságot vagy kapcsolatot fejez ki.

prima facie - első ránézésre.

prímszámolyan szám, amelynek egyetlen tényezője önmaga és az egy. **programaz a** szabály, amely egy regisztergép műveleteit szabályozza. **qualiaa** a szubjektív, kvalitatív szakfilozófiai kifejezés,

kiváltságos, első személyű tapasztalati aspektusok.

rekurzív definíció egy végtelen osztály véges specifikációjának eszköze, amely egy alaptételből és egy rekurzív tételből áll. A bináris számok rekurzívan a következőképpen definiálhatók: az egy és a nulla egyaránt bináris szám (alaptétel), és bármely bináris számhoz egy egyest vagy egy nullát hozzáadva bináris számot kapunk (rekurzív tétel).

reflex ív pavlovi kifejezés, amely az inger és a viselkedés közötti kapcsolatot hivatott magyarázni.

regiszter gép egyfajta *formális rendszer*, amely használható a következők meghatározására
számítás.

a *formális rendszert* alkotó két gyűjtemény egyikének **szabályai**. Bármely adott *szabály* és bármely adott *állapot* esetén *effektívnek* kell lennie, hogy a szabály érvényes-e vagy sem, és ha igen, akkor *effektív módon* véges számú kimeneti állapotot kell szolgáltatnia.

szemantika tágabb értelemben, jelentése.

szonoráns olyan *fonéma*, amelynek előállításához a levegőnek az artikulációs készüléken való áthaladása nem akadályozható.

a *formális rendszert* alkotó két gyűjtemény egyikét **jelenti**. *Állapotokat* az entitások bármely gyűjteménye fölött meg lehet

határozni, feltéve, hogy van *effektív* eljárás bármely két állapot közötti különbségtételre.

szalmabábu tévedésegy ellentétes álláspontot gyengébbnek minősíteni, mint amilyen az valójában, majd a gyengébb álláspont ellen érvelni. **szubsztrátfüggetlenség** a képesség, hogy bármilyen szubsztrátban megvalósuljon - az elme funkcionalista elméleteinek tulajdonsága, hogy a mentális állapotok *szubsztrátfüggetlennek* minősül.

szimmetrikus a relációk olyan tulajdonsága, hogy minden A és B esetében, ha A a B-vel való relációban áll, akkor B az A-val való relációban áll.

szinapszis *neuronok* közötti kapcsolat.

szintaxis a szabály által szabályozott szimbólumkombináció.

végállapot egy *formális rendszer generált állapota*, amely nem tartalmaz olyan

a rendszer *szabályai* vonatkoznak rá.

token-fizikalizmus az elme filozófiai elmélete, amely szerint, amikor egy alany mentális állapotban van, akkor valamilyen neurális állapotban van, de nem történik azonosítás a mentális és neurális állapotok típusai között.

transzdukció az a folyamat, amelynek során egy elektromos jel átalakul kémiai jellé és *fordítva*.

átviteli függvény a *mesterséges neurális* hálózat minden egyes termelő csomópontja által végrehajtott két függvény egyike, amely meghatározza, hogy egy csomópont aktiválási szintje milyen legyen az adott csomópont eltérő aktiválási szintje (és esetleg az azt megelőző aktiválási szintje) mellett.

tranzitív a relációk olyan tulajdonsága, hogy minden A, B és C esetében, ha A a B-vel való relációban áll, és B a C-vel való relációban áll, akkor A a C-vel való relációban áll.

A Turing-teszt egy olyan teszt, amelyben emberi és számítógépes résztvevők és egy emberi kérdező vesznek részt, és ha a számítógép képes megtéveszteni az emberi kérdezőt, hogy azt higgye, hogy ember, akkor azt kell mondanunk a számítógépről, hogy van elméje.

univerzális gép a regisztergépek egy speciális fajtája, amely a *Gödel-kódolásnak köszönhetően* képes bármilyen regisztergép-programot bemenetként fogadni és működtetni.

érvényesség a logikai következtetési formák olyan tulajdonsága, hogy a premissák igazsága garantálja a következtetés igazságát - az *érvényes* következtetések nem engednek meg *ellenpéldákat*.

A wetware az agyra, mint információfeldolgozó hardverre utal.

INDEX

- hiányzó qualia *lásd*
- zombik aktiválása, 188,
190
- aktiválási függvény, 188-9, 200
- adicity, 77
- algoritmusok *lásd* hatékony
eljárások allofonok, 170-1
kétértelműsé
 g lexikális,
 153
 szintaktikus *lásd*
- amfiboly amfiboly, 153
- amygdala, 30
- anomális monizmus *lásd* tulajdon
 dualizmus
- antecedens (a feltételes módban),
134 arcuate fasciculus, 32
- Armstrong, D., 36
- mesterséges intelligencia, 100
 mesterséges neurális hálózatok,
 187-201 tréning *lásd*
 backpropagation ausztrál
 materializmus, 35, 37-42
- axon, 33
- backpropagation, 197-8
- bazális ganglionok, 30
- a kérdés feltevése, 39, 50-1
- viselkedéses kondicionálás
lásd
 kondicionáló
- behaviorizmus
 analitikus *lásd* filozófiai
 módszertani *lásd*
- pszichológiai

filozófiai, 23-6, 35, 45
 pszichológiai, 20-2
 elágazási tényező, 114
 Broca, P., 31
 Broca-féle afázia, 31
 Broca-terület, 31

 karburátor, 44-5
 kartéziánus dualizmus *lásd*
 dualizmus kategorikus
 gondolkodás, 158
 ok-okozati zárás, 11
 kauzális elmélet (az elméről), 35-7,
 38,
 45
 központi állami materializmus *lásd*
 ausztrál materializmus kisagy,
 28
 Chalmers, D., 203
 sakk, 53-6, 104, 128-31
 Kínai szoba, 176-9
 Chomsky, N., 149, 164
 Church, A., 87
 Church/Turing tézis, 87-8, 93,
 200
 kognitív architektúra, 185-6;
 lásd még mesterséges
 neurális hálózatok
 kompozicionalitás, 182
 kiszámíthatóság, 77-8, 87-8, 200
 számítás, 75-8, 95
 komputacionalizmus, 95-108
 számítógép, 93
 Comte, A., 20

- feltételeesség, 134
kondicionálás
 operáns, 21-2
 Pavlovian, 20, 2
vezetési afázia, 32
kötés, 136
konjunktusok, 137
konnekcionizmus, 185-6; *lásd még*
 mesterséges neurális
hálózatok tudatossága, 202-3
következetes (a feltételes), 134
tartalom, 102-3, 175; *lásd még*
 szemantika
kontextusérzékenység, 191-6
corpus callosum, 30
corvée, 1
ellenpéldák, 156, 160
kreativitás, 106-8
kereszt-lateralizáció, 30-1
Cyc/Cycorp, 144
- eldönthetőség, 86
döntési probléma *lásd*
eldönthetőségi következtetés,
133-7
dendritek, 33
levezetés, 62-4, 66, 67
Descartes, R., 5, 29
diftongok *lásd* szonoránsok
diszpozíciók, 23-6
elosztott képviselő *lásd*
 dopamin képviselő, 28
kettős aspektus elmélet *lásd a*
 tulajdon dualizmus
kartéziánus
 dualizmus, 5-
 12
 epifenomenalizmus, 13-14
 interakcionista *lásd* karteziánus
 occasionalizmus, 13
 párhuzamosság, 12-13
 ingatlan, 14
 anyag, 4-5
- Ebbinghaus, H., 18, 19
hatékony eljárások, 53-7, 64, 67,
 71-3, 76, 86-8, 97, 200
megtestesülés, 179-80, 206
érzelmeik, 30, 205-6
endokrin mirigyek, 29
Következtetés *lásd* logikai
következmény
Entscheidungsproblem *lásd* döntés
 problémás
epifenomén, 14
epifenoméanalizmus *lásd*
dualizmus lényegi
tulajdonságok, 8-9
kizárólagos diszjunkció, 190
szakértői rendszerek, 137-44
exponenciális, 89
- Fechner, G., 16-17
Fibonacci-sorozat, 84
formális rendszerek, 57-69, 71-2, 88, 95,
 102
homloklebeny, 31
függvény (matematikai) *lásd*
 totális/részleges
funkciók funkcionalizmus,
44-51, 95, 101
- a szerencsejátékos tévedése, 158
generáció, 61-2
generációs fák, 64-7
generatív nyelvtan
 lásd nyelvtan
szellem a gépben, 23, 26
Isten, 12-13
Gödel, K., 88
Gödel-kódolás, 88-
94 nyelvtan
 generatív, 149-52
 mondatszerkezet, 150
- megállási probléma, 76, 85, 86
Helmholtz, H., 16-17
heurisztikus funkció, 117, 125-6,
128

heurisztikus

keresés a

legjobb első,

120-1

hegymászás, 118-20

hippokampusz, 30

homonimák, 148

homofonok, 148

Huntington-kör, 30

hipotalamusz, 29

identitás *lásd*: személyes

identitás identitás (elme)

identitáselmélet *lásd*:

identitáselmélet (tudat)

ausztrál materializmus

imitációs játék *lásd* Turing-

teszt imitációs ember *lásd*

zombik indukció, 133

intencionalitás, 181-2, 203

interakcionista dualizmus *lásd*

dualizmus introspekció, 7, 17-20,
38-9

fordított spektrum, 48-9

irracionalitás *lásd* racionalitás

izomorfizmus, 68, 102-3, 104

Jackson, F., 43

Johnson-Laird, P., 161

Kahneman, D., 161

Külpe, O., 18-19

laterális genikuláris mag, 29

tanulási algoritmusok, 105-6

Leibniz törvénye, 39

limbikus rendszer, 30

nyelvi kapacitás, 109

Loebner-díj, 111

logikai kapuk, 189-91, 200

logika/logikák, 132-3

logikai következmény, 132; *lásd*
még

érvényesség

logikai forma, 135-

Mach, E., 20

materializmus

Ausztrál *lásd* ausztrál
 materializmus
 központi állam *lásd*
 ausztrál
 materializmus
 reduktív *lásd*
 ausztrál
 materializmus
 jelentés *lásd* szemantika
 medialis geniculáris mag,
 29 melatonin, 29
 memória, 18, 30, 203
 mentális modellek,
 160-1 mentális
 reprezentáció *lásd*
 képviselési
 minimax, 125-6
 modus ponens, 134
 modus tollens, 157
 monophongok *lásd*
 szonoránsok többszörös
 megvalósíthatóság, 40-1,
 101
 MYCIN, 143
 myelin, 33

 neurális biomérnöki tevékenység,
 200-1
 neurális plaszticitás, 32, 41
 neuronok, 32-4
 neurotranszmitterek, 28, 29, 200,
 205

 obstruensek, 165-7; *lásd még*
 fonémák
 occasionalizmus *lásd*
 dualizmus occipitális
 lebeny, 32
 Ockham borotvája, 10-11, 24, 38
 szaglógumók, 29-30
 ontológia, 4
 operáns kondicionálás *lásd*
 kondicionáló
 operációs rendszer,
 104-5

túlextenzió (szemantikai), 172
 oxitocin, 29

paradigmák *lásd*
 sztereotípiák
 párhuzamosság *lásd*
 dualizmus

fali lebeny, 31
 Parkinson-kór, 28, 30
 mintakivonatolás, 131, 163, 198-9
 Pavlov, I., 20
 Pavlovi kondicionálás *lásd*
 a személyes
 identitás
 kondicionálása, 203-5
 petitio principii *lásd* a kérdés
 felvetése
 fonémák, 146, 165-73
 fonetikai tulajdonságok, 145; *lásd*
még
 fonetikai realizációs
 szabályok fonetikai realizációs
 szabályok, 170-2 fonológiai
 egyszerűsítés, 173
 fonológia, 164-73
 mondatszerkezeti nyelvtan *lásd*
 nyelvtan
 mondatszerkezeti fák, 150-2;
 lásd még: generációs fák
 tobozmirigy, 29
 agyalapi mirigy, 29
 Place, U. T., 36, 37
 pons, 27-8
 pozitívizmus, 20, 22
 predikátumok *lásd*
 predikáció predikáció,
 134-7 eleve elrendelt
 harmónia *lásd*
 párhuzamosság
 elsődleges hallókéreg, 31
 elsődleges motoros kéreg, 31
 elsődleges érzékszalag, 31
 elsődleges látókéreg, 32
 prímuszámok, 89 produktivitás
 (a nyelv), 149 program, 71-5,
 79-84, 90-3, 96
 bizonyítás *lásd* dedukció,
 levezetés, indukció
 tulajdon dualizmus *lásd*
 dualizmus metszés, 127-8
 pszichológia
 behaviorista, 20-2

- introspekcionista, 17-20
 élettani, 16-17
- qualia, 42-3, 48-51, 203
- racionalitás, 162-3
- Rayner, R., 21
- rekurzív
- meghatározás, 59
- reduktív
- materializmus
- lásd*
- ausztrál
- materializmus
- reflexívek, 20-1
- nyilvántartó gépek, 70-6, 78-84,
 86-7, 88, 89-93, 96
- szimmetrikus
- kapcsolato
- k, 137
- tranzitív, 136
- képviselő, 180, 181-5
- elosztva, 184-5
- szimbolikus, 183-4
- robotok, 1
- Ryle, G., 23
- elégedettség,
- 135, 137
- keresés
- (vakon)
- alulról felfelé, 114, 142-3
- először a szélesség, 115, 117
- először a mélység, 116-17
- heurisztikus *lásd*
- heurisztikus keresés
- felülről lefelé, 113-
- 14
- Searle, J., 176
- szemantika, 146-8,
- 175-80 Skinner, B.
- F., 21-2 Smart, J.
- J. C., 36, 37
- soma, 33
- szomatosenzoros kéreg
- lásd* elsődleges
- érzékelő sáv
- szonoránsok, 167-70; *lásd még*
- fonémák
- lélek, 6
- faji sovinizmus, 41, 47
- beszédszintézis, 191-6, 199
- négyzetelés, 81-3; *lásd még*
- exponenciálás

- sztereotípiák, 158, 159; *lásd még*
 mentális
 modellek inger-
 válasz, 17
 szalmabábu, 94
 karakterláncok összekapcsolása,
 58-9
 string változók, 58
 szubsztancia dualizmus *lásd*
 dualizmus substantia nigra, 28
 szubsztrátfüggetlenség, 47, 101
 szimbolikus reprezentáció *ld.*
 képviselő
 szimmetrikus kapcsolat *lásd*
 kapcsolatok szinapszis, 33
 szinaptikus hasadék, 33
 szinaptikus vezikulák, 33
 szintaxis, 147-9, 175-9; *lásd még*
 nyelvtan
- halántéklebény, 31-2
 thalamus, 28-9, 30
 termodinamika, 11
 token identitáselmélet (az elme)
lásd
 jelképes fizikalizmus
 jelképes fizikalizmus, 41-2, 46-7,
 101
 topológiai szerveződés, 31
 teljes/részleges funkciók, 77
- transzdukció, 33
 transzferfüggvény, 188-9, 200
 tranzitív reláció *lásd* relációk
 Turing, A., 86, 109-10
 Turing-gépek, 86-7; *lásd még*
 regisztrálógépek
 Turing-teszt, 109-11
 Tversky, A., 161
 típus/token azonosság, 46
- alulbővítés (szemantikai), 172-3
 univerzális gépek, 92-3, 96-7, 98
- érvényesség, 99, 156, 157
 magánhangzók, kardinális, 168; *lásd*
még
 szonoránsok
- Watson, J., 20-1
 Wernicke, K., 32
 Wernicke afázia, 32
 Wernicke területe, 31-2
 wetware, 99, 101
Wizard Of Oz, The, 148
 Wundt, W., 17-19
- zombik, 50-1