

Nick Boström - Kas me elame arvutisimulatsioonis?

Urmas Pitsi, märts 2020

Kokkuvõte

Boströmi argumendi kohaselt on *vähemalt üks* järgnevatest väidetest tõene:

1. Inimkond sureb välja enne, kui jõuab “posthuman” arengutasemeni;
2. On äärmiselt ebatõenäoline, et ükski posthuman tsivilisatsioon simuleerib enda evolutsioonilist ajalugu (või mõnd selle variatsiooni);
3. Me elame paaegu kindlasti arvutisimulatsioonis.

Sellest tulenevalt on uskumus sellest, et on märkimisväärne võimalus posthuman-itest, kes teevad eellaste simulatsioone on väär, välja arvatud juhul, kui me juba praegu elame arvutisimulatsioonis. Lisaks veel mitmed muud järeldused, mida arutleme allpool.

Boströmi artikkel on jaotatud 7-ks paragrahviks:

- I. Sissejuhatus
- II. Eeldus
- III. Arvutatavuse tehnoloogilised piirid
- IV. Simulatsiooniargumendi tuum
- V. Ükskõiksusprintsip (võrdse tõenäosuse printsip)
- VI. Tõlgendus
- VII. Järeldus

I. Sissejuhatus

Struktuur on järgnev:

- (1) Formuleerime eelduse, mille me võtame mõistuse filosoofiast (philosophy of mind);
- (2) Vaatleme kogemuslikult millistel alustel me saame mõelda, et tuleviku inimkond on võimeline simuleerima inim mõistuseid. Arvestades, et see oleks kooskõlas meile teadaolevate füüsikaseadustega ja inseneeria piirangutega. See osa ei ole filosoofiliselt vajalik, küll aga pakub motivatsiooni järgnevat tähelepanelikumalt jälgima.
- (3) Argumendi tuum: kasutab lihtsat tõenäosusteooriat ja ükskõiksuse printsiipi. Ükskõiksuse printsiip: (võrdse tõenäosuse printsiip, principle of indifference) tõenduse puudumisel jaotame sündmuste toimumiste tõenäosused võrdselt kõigi võimaluste vahel.
- (4) Lõpetuseks arutleme tulemuseks saadud 3-väitelise järelduse tõlgendusi.

II. Eeldus

- Aineest-sõltumatus (*substrate-independence*)

Tavaline eeldus mõistusefilosoofias on *ainest-sõltumatuse* (*substrate-independence*) eeldus. Idee seisneb selles, et **mentaalsed olekud on sõltumatud füüsilisest aineest**, milles need parasjagu tekivad. Sellest tulenevalt, saab inimaju tegevust modelleerida arvutiga, teostades korrektsed arvutuslikud struktuurid ja protsessid, mis on seotud teadvusega (teadlike kogemustega). Arvuti ei pea olema ilmtingimata moodustatud bioloogilistest materjalidest, vaid võib ka olla muust materjalist, peaasi, et suudab teostada samasuguseid tegevusi, nagu seda teeb inimaju.

- Aineest-sõltumatus pole vajalik

Aineest-sõltumatuse eeldus ei ole range, st see pole *vajalik*. Piisab, et arvutusi teostav masin omaks teadvust. Veel enam, mõistuse loomiseks pole vaja isegi inimekäitumise täielikku imiteerimist. Meil on vaja täita ainult nõrgem eeldus, mis seisneb järgnevas. **Piisab subjektiivsete kogemuste genereerimisest, nii et inimaju protsessid on taasesitatud piisava täpsusega**, näiteks üksikute sünapside (aju-närvirakkude vahelised ühendused) tasemel. Taoline leebem versioon aineest-sõltumatusest on üsna laialt aktsepteeritud.

Ajus on ka sünapsidest väiksemal skaalal osakesi, aga me **eeldame, et suudame modelleerida subjektiivseid kogemusi sünapside tasandil**.

III. Arvutatavuse tehnoloogilised piirid

- **Hetkel piisav arvutusvõimekus puudub, aga me saavutame selle (varsti)**

Käesoleval ajal puudub riistvara ja tarkvara, mille abil saaks arvutites luua teadvust. Meil on veenvaid argumente, et *kui* tehnoloogia areng jätkub samas tempos, *siis* me ületame selle takistuse.

- **Pole ajalisi piiranguid: “posthuman” tasemeni võib kuluda ka sadu tuhandeid aastaid**

- **Inimkond omandab enamiku tehnoloogilisest võimekusest, mis on vastavuses füüsika seadustega ning materjali- ja energiapiirangutega**

Kuna meil hetkel puudub “kõige teooria” (theory of everything), siis ei saa välistada uusi füüsikalisi nähtusi, mis praeguses füüsikateoorias on võimatud. Näiteks Bremermann-Bekensteini piir ja Musta Augu limiit (? U.P.: kas see pole eelmise lausega otseses vastuolus?).

- **Mõned hetkel teadaolevad füüsikalised piirid**

Bremermanni limiit [5]: arvutusvõimsuse maksimummäär, mida saab välja “pigistada” 1kg mateeriat meie Universumis. $\approx 1.36 \times 10^{50}$ **bitti sekundis kilogrammi kohta**.

Beckensteini piir [6]: kui palju informatsiooni saab **maksimaalselt salvestada** ühte ruumiossa $\approx 2.58 \times 10^{43}$ **bitti kilogrammi ja meetri kohta** (meeter on siin vaadeldava ruumi raadius).

Erinevad autorid on arvutanud välja erinevaid arvutusvõimsuse hinnanguid.

Maksimum kiirus, Lloyd [7]: 5×10^{50} **loogilist operatsiooni sekundis**

- **Boströmi järeldus: inimaju simuleerimiseks piisab arvutuskiirusest $\sim 10^{16} - 10^{17}$ operatsiooni sekundis**

Tuletuskäik: Neuronite arv * sünapsid neuroni kohta * 10 korda sekundis: $\sim 10^{10} \times 10^5 \times 10$

- **Arvutimängu analoog: kuvatakse seda osa pildist, mida vaadatakse**

Näiteks Maakera sisemust polegi vaja kunagi välja joonistada, ega ka rakkusid ja kosmost. Piisab hägusest pildist. Vajadusel saab simuleerija teha ka parandusi ja simulatsiooni tagasi kerida, näiteks mõne vastuolu ilmnemisel. Simulatsiooni juhendaja saab vastuolulisi kohti vahele jätta. Seega on igati tõenäoline, et inimõistuse simuleerimiseks piisab neuronite või neuronite sisese taseme simuleerimisest.

- **Hinnanguline realistliku inimajaloo simulatsiooni kulu on seega $\sim 10^{33} - 10^{36}$ operatsiooni**
100 milardit inimest * 50 aastat/inimene * 30 milj sek/aastas * $[10^{14}, 10^{17}]$ operatsiooni igas inimajus
sekundis $\approx 10^{33} - 10^{36}$.

- **Teadaoleva nanotehnoloogia baasil saab luua planeedi suuruse arvuti, mis teeb 10^{42} operatsiooni sekundis ning võib eeldada, et see on kaugel optimaalsest [8]**

Üks taoline arvuti suudab teostada kogu inimkonna ajaloo simulatsiooni, kasutades vähem kui miljondikku oma arvutusvõimsusest ühes sekundis. Tuleviku inimesed võivad ehitada astronoomilise hulga selliseid arvuteid.

- **Järeldus: Tulevikuinimestel on piisavalt arvutusvõimsust, et teostada ülilpalju eellaste simulatsioone, kasutades seejuures vaid tibatillukest osa saadaolevast arvutusvõimsusest**

IV. Simulatsiooniargumendi tuum

- Kui on väga tõenäoline, et meie tsivilisatsioon areneb posthuman tasemele ja teostab palju eellaste simulatsioone, siis võiks järeldada, et ka meie elame juba ühes sellises simulatsioonis

f_p : inim-tasemel tehnoloogilisel arengutasemel tsivilisatsioonide osakaal, mis jõuavad posthuman tasemeni.

\overline{N} : Keskmine arv eellaste simulatsioone, mida teostab üks posthuman tsivilisatsioon.

\overline{H} : Keskmine arv indiviide, kes on elanud tsivilisatsioonis enne, kui see jõuab posthuman tasemeni.

Tegelik osakaal kõikidest inimtüüpi kogemustega jälgijatest, kes elavad simulatsioonides:

$$f_{sim} = \frac{f_p \overline{N} \overline{H}}{f_p \overline{N} \overline{H} + \overline{H}}$$

Olgu:

f_i : eellaste simulatsioonidest huvitatud tsivilisatsioonide osakaal.

\overline{Ni} : keskmine arv eellaste simulatsioone mida huvitatud tsivilisatsioonid teostavad.

Saame:

$$\overline{N} = f_i \overline{Ni}$$

Seega:

$$f_{sim} = \frac{f_p f_i \overline{Ni}}{f_p f_i \overline{Ni} + 1} \quad (*)$$

Kuna posthuman tsivilisatsioonil on erakordselt võimsad arvutid, siis ni on väga suur, nagu me nägime eelmises peatükis. Vaadeldes (*) võime näha, et *vähemalt üks* järgnevast kolmest väitest peab olema tõene:

$$(1) f_p \approx 0$$

$$(2) f_i \approx 0$$

$$(3) f_{sim} \approx 1$$

V. Ükskõiksusprintsip (võrdse tõenäosuse printsip)

Kuna meil puuduvad selged tõendid, kuidas jaotada tõenäosusi alternatiivide vahel, siis jaotame sündmuste toimumiste tõenäosused võrdselt kõigi võimaluste vahel.

Argumendi 3 osa saavad igaüks tõenäosuseks $1/3$.

VI. Tõlgendus

- **Kui (1) on tõene, siis inimkon(na)d ei jõua posthuman tasemeni**
Seega on üsna keeruline õigustada miks peaks meie tsivilisatsioon olema privilegeeritud seisuses ja vältides katastroofe jõudma posthuman tasemeni. Inimkonna hukkumise tõenäosus ~100%. Alternatiivselt võib olla, et madalama taseme tsivilisatsioonid jäävad Maal ellu alati, samas kui teatud arengutasemeni jõudvad tsivilisatsioonid hukkuvad.
- **Alternatiivi (2) tõesuse korral** peaks eksisteerima mingi veenev põhjus, miks kaugelearenenud tsivilisatsioonid ei tahaks simuleerida maailmasid: jõukus hajutatud, kellelgi üksikuna pole ressursi; simulatsioon ei anna uut informatsiooni, seetõttu ebahuvitav; moraalsed põhjused.
- **Alternatiiv (3)** on kõige intrigeerivam. Maailm, mida me näeme, pole “reaalne”, see pole reaalsuse algtasand. Sellised simulatsioon-maalilmad on nagu “virtuaalmasinad” arvutimaailmas, kus üks programm jookseb teise - simuleeritud - arvuti sees. Niimoodi võib simulatsioone aheldada lõpmatult üksteise sisse. Kui me suudame luua eellaste simulatsioone, siis see oleks tugev tõendus (1) ja (2) vastu, millest peaksime järeldama, et me elame simulatsioonis. Veel enam, peaksime kahtlustama, et meie simuleerijad on ise ka simuleeritud olevused; ja nende loojad omakorda võivad samuti olla simuleeritud. Reaalsusel on seega mitu tasandit. Tasandite arvu võib piirata ressursside piiratud maht. Näiteks võime eeldada, et meie simulatsioon võidaks peatada kui saavutame posthuman taseme.
- **Eeldades, et elame simulatsioonis, millised võiksid olla järeldused meile, inimestele?**
Jätkame elu ja planeerimist, nagu oleme siiani teinud. See tundub kõige mõistlikum alternatiiv.

VII. Järeldus

- Tehnoloogiliselt edasiarenenud (mature) “posthuman” tsivilisatsioonil on kasutada ülimalt suur arvutusvõimsus. Põhinedes sellele kogemuslikule faktile, näitab simulatsiooniargument, et *vähemalt üks järgnevatest väidetest on tõene*:
 - (1) “Posthuman” tasemeni arenenud inimtsivilisatsioonide osakaal on ligikaudu null.
 - (2) “Posthuman” tasemeni arenenud ja eellaste simulatsioonist huvitatud inimtsivilisatsioonide osakaal on ligikaudu null.
 - (3) Meiesarnaste kogemustega inimeste osakaal, kes elavad simulatsioonis on ligikaudu üks.
- Kui (1) on tõene, siis peaaegu kindlasti me sureme välja enne “posthuman” tasemeni jõudmist.
- Kui (2) on tõene, siis peavad edasijõudnud tsivilisatsioonid tugevalt koonduma niimoodi, et neis poleks ühtegi suhteliselt jõukat indiviidi, kes sooviks teostada eellaste simulatsiooni.
- Kui (3) on tõene, siis me elame peaaegu kindlasti simulatsioonis.
- Kuna meil puudub täpsem informatsioon, siis on mõistlik **jaotada hinnangulised tõenäosused võrdselt (1), (2) ja (3) vahel**.
- Kui on täidetud tingimus, et me praegu ei ela simulatsioonis, siis meie järeltulijad ei teosta peaaegu kindlasti mitte kunagi eellaste simulatsiooni.