

Vad är egentligen poängen med det så kallade "Turing-testet"? Att avgöra om en maskin kan tänka? Ibland presenteras testet på det sättet. Och då finns det förstås en massa invändningar att rikta mot det, både från den som är entusiastisk inför idén om tänkande maskiner och från den som är skeptisk.

Entusiasten kan undra varför tänkande nödvändigtvis måste handla om att kunna imitera en människa. När allt kommer omkring är ju vi människor ganska så begränsade i vårt tänkande och gör allehanda misstag, medan idén om en tänkande maskin rymmer förhoppningen om en mer fulländad intelligens. En maskin som tänker perfekt skulle snart avslöjas i det imitationsspel som Turing-testet iscensätter. För att konstruera en dator som klarar testet måste vi alltså, tycks det, bygga in en viss benägenhet till logiska kullerbyttor och andra tankemässiga ofullkomligheter. Men varför skulle *det* vara ett led i att få maskinen att tänka?

Skeptikern kan å sin sida invända att testet bygger på en väldigt fyrkantig idé om tänkande. Varför skulle några minuters konversation där de två konversationspartnerna är dolda för varandra av en skärm, göra rättvisa åt vad tänkande innebär? Ja, finns det inte rentav en mycket tvivelaktig dualism inbyggd i testets själva design? Varför skulle det som skärmen döljer – min konversationspartners kropp, gester, minspel och icke-verbala praktiska förmågor – vara oväsentligt för frågan om det är en tänkande varelse jag har framför mig? Filosofer har sedan länge ifrågasatt idén att tänkande är frikopplat från kroppslighet i denna konkreta mening. Men många av Turing-testets förespråkare tycks helt enkelt ta för givet att skärmen är nödvändig för att isolera det fenomen man vill studera – "tänkandet i sig". Vad det nu skulle vara.

Nå; hur är det då med testets ursprunglige konstruktör, Alan Turing själv? Vad var egentligen hans tanke med testet? Går man till källorna – dvs. den artikel där han presenterade testet, "Computing Machinery and Intelligence" från 1950 – så framträder en rätt överraskande och intressant bild.

Turing är i själva verket mycket noggrann med att påpeka att syftet med testet *inte* är att avgöra om en maskin kan tänka! Han anser nämligen att frågan "Kan maskiner tänka?" är nonsens: "The original question 'Can machines think?' I believe to be too meaningless to deserve discussion" (s. 442). Istället är hans mål att *ersätta* denna meningslösa fråga med en som har klar innebörd: Kan vi konstruera en maskin som klarar att imitera en människa på det sätt Turing-testet beskriver? Sedan parar Turing denna ambition med en *förutsägelse*: Om 50 år (dvs. vid sekelskiftet 2000) kommer man att ha lyckats konstruera maskiner som klarar testet, och om man då säger om dessa maskiner att de "tänker" så kommer man inte att bli motsagd. Med Turings egna ord: "I believe that at the end of the century the use of words and general educated opinion will have altered so much that one will be able to speak of machines thinking without expecting to be contradicted" (s. 442).

Här vill någon kanske hävda att Turing rundar den verkligt viktiga frågan, nämligen om en sådan framtida förändring av språkanvändningen vore *berättigad* eller ej – kan en maskin som klarar Turing-testet *verkligt* tänka? Men Turings poäng är ju att denna "verkligt viktiga fråga" inte är någon begriplig fråga alls, och att vi kan få till en vettig diskussion bara om vi lägger den åt sidan.

Turings ståndpunkt kan tyckas märklig. För den fråga han utrangerar som obegriplig innehåller ju idel välbekanta ord – "kan", "maskiner" och "tänka". Dessa ord lär vi oss använda och förstå redan som barn. Måste då inte sammansättningen av dem vara begriplig? Nej, menar Turing – och här tror jag att han har rätt. Frågor kan vara meningslösa även om de innehåller välbekanta och i andra sammanhang begripliga ord. Och den formen av meningslöshet är särskilt lurig, just eftersom de ingående orden tycks så uppenbart meningsfulla.

Sedan är det en annan sak om Turing verkligen har rätt i att just denna fråga – "Kan maskiner tänka?" – är meningslös. Jag har inte utrymme att redan ut den saken här. Däremot skulle jag vilja peka på vad som blir konsekvensen av att transformera problematiken på det sätt Turing rekommenderar. Jag tror nämligen att inte så få

vetenskapsmän och filosofer delar Turings grundläggande attityd, att det lämpliga sättet att närma sig djupa filosofiska frågor är att ersätta dem med andra frågor som saknar det filosofiska djupet men som går att angripa på ett vetenskapligt sätt. Men sedan frestas man ändå att presentera saken som om den ursprungliga frågan trots allt får ett svar genom de vetenskapliga resultat man når. Innebär inte ett sådant tillvägagångssätt att det som verkligen förundrade eller bekymrade oss när vi funderade kring frågan sopas under mattan, medan vi luras att tro att frågan besvarats?

Om påståendet att det är möjligt att konstruera en maskin som kan tänka ersätts med påståendet att vi kommer att kunna konstruera en maskin som klarar Turingtestet, parat med förutsägelsen att våra språk- och tankevanor då kommer att ha förändrats så att vi inte kommer att invända mot någon som beskriver denna maskin som "tänkande" – vad innebär detta för själva *diskussionen* av frågan om maskiner kan tänka?

Jag menar: Vad innebär det i så fall att vara *entusiastisk* inför idén att maskiner kan tänka? Och vad innebär det att vara *skeptisk* inför samma idé? Att helt enkelt tro eller inte tro på förutsägelsen? Knappast. Entusiasmen innebär ju också att man *hoppas* på att den förutsagda förändringen av våra språk- och tankevanor ska inträffa, medan skepticismen innebär att man hoppas på att förändringen *inte* kommer att inträffa. Men i så fall verkar motsättningen mellan entusiasten och skeptikern inte längre anta formen av en rent filosofisk eller vetenskaplig diskussion, utan nu få en tydligt moralisk-politisk karaktär, där det handlar om att hitta medel att påskynda respektive förhindra den förutsagda förändringen – medel som inte nödvändigtvis behöver handla om goda argument, utan mer om rent retoriska övertalningsmedel. Entusiasten kommer att leta efter retoriska vägar att övertyga folk om att börja tala om maskiner som "tänkande", medan skeptikern kommer att saluföra en motsatt form av retorik som förhindrar en sådan förändring.

Och är det inte något sådant vi faktiskt ser hända i mycket av den samtida diskussionen kring om maskiner kan tänka eller inte? Tänk på alla häftiga videosnuttar man kan hitta på YouTube om man söker på "AGI", "robotics" osv.; tänk på hur forskare aktivt bearbetar tidnings- och TV-redaktioner för att de ska presentera spektakulära övertolkningar av forskningsresultat som vore de etablerade sanningar. Och tänk på hur skeptiker i det motsatta lägret opponerar sig mot tanken på tänkande maskiner utan att egentligen klargöra vad som skulle vara förvirrat med denna idé.

Turing ville naturligtvis inte själv att debatten skulle anta denna egendomliga karaktär; han syfte var förstås att vara så klarsynt som möjligt när det gäller problemet kring tänkande maskiner. Dock tycks det mig som om hans angreppssätt innebär att det blir svårt att värja sig mot just en sådan utveckling, "retorisering", av debatten – för vad kan Turing egentligen ha att invända mot att entusiaster försöker påverka våra språk- och tankevanor så att hans förutsägelse går i uppfyllelse? Som kallt betraktande vetenskapsman har han väl bara att konstatera att utvecklingen går i den riktning han förutsagt. Att ställa frågan om denna utveckling är vetenskapligt eller filosofiskt berättigad, har han ju utdömt som meningslös.

Martin Gustafsson