

## Varför Turingtestet är vetenskapligt ointressant

Nils Dahlbäck  
Linköpings universitet  
nils.dahlback@liu.se

### Inledning

Turingtestet anses vara *one of the philosophical foundations of Artificial intelligence* (Proudfoot (2013) p 391). Jag skall här försöka visa varför jag inte delar den uppfattningen, utan istället anser att det är en återvändsgränd som alltför många har irrat runt i alltför länge.

Mina kritiska punkter rör två områden, metodologiska och begreppsmässiga.

Turing var en brilliant matematiker och logiker utan djupa kunskaper i beteendevetenskaplig teori och metod. Det är fullt begripligt att han inte kunde prestera bättre än vad han gjorde. Nu har dessutom både beteendevetenskaplig teori och metod utvecklats betydligt sedan Turing publicerade sitt papper, vilket gör att bristerna blir än mer uppenbara, men ingen kommentator av pappret som jag känner till verkar ha tagit detta i beaktande, utan behandlat pappret som om det vore invändningsfritt i dessa avseenden.

### Metodologiska problem

Om vi accepterar Proudfoots tolkning så verkar det handla om ett test om huruvida människor kan fås att tro att en korrekt programmerad dator är en människa.

Antingen vill man göra detta för att testa den omvända tesen, dvs. att det inte finns någon människa (i den grupp som Turing beskriver) som kan fås att tro att en dator är en människa. I så fall räcker det förstås med att finna en person som i användning, eller kanske snarare i dialog med en dator, tror att det är en människa. Men att det finns individer som tror det har vi vetat sedan ELIZA (Weizenbaum, 1966). Så det testet behövs inte. Det är redan gjort.

Men skall vi istället undersöka om detta gäller för en den grupp som Turing beskriver, eller i alla fall för majoriteten av dessa, så räcker det med att ha läst en introduktionskurs i beteendevetenskaplig metod för att veta att det inte räcker med att testa en enda person. Det behövs ett stort sampel ur den relevanta populationen. Och ändå skriver allt kommentatorer jag känner till som om det handlade om en och endast en person.

Sedan finns det ytterligare uppenbara problem. Hur länge skall dialogen få pågå? Vi vet redan från ELIZA att ju längre tiden går blir antalet som inser att det inte är en människa (eller möjligen att det är en synnerligen retsam och otrevlig människa) som man "talar" med. Här har vad jag vet ingen kommentator lagt fram några förslag till avgränsningar, men samtidigt är det uppenbart att det inte kan handla om oändlig tid.

Dessutom finns möjligen här också det problem som Herb Clark tar upp i sitt 1973-papper "The Fixed-effect Fallacy", alltså att man, eftersom man inte kan ta med alla potentiella yttranden, eller dialoger, eller vilken enhet man nu väljer att betrakta som relevant, utan man måste göra ett urval. I så fall måste man göra den statistiska testningen så att inte bara individerna som deltar utan också yttrandena som görs är urval ur ett sampel, vilket i sin tur kräver att man gör ett s.k. F' test (vilket gör det betydligt svårare att uppnå statistik signifikans).

Vad jag kan se gäller samtliga invändningar ovan också för alternativa tolkningar av vad Turingtestet mäter som om det är ett test på om datorer kan tänka eller vara intelligenta eller något annat.

MANUS UNDER UTARBETANDE. CITERA EJ UTAN SKRIFTLIGT MEDIGIVANDE FRÅN FÖRFATTAREN.  
Men kommentera gärna!

Slutsatsen är uppenbar; skall Turintestet gå från att vara ett tankeexperiment till ett faktiskt utfört experiment måste metodiken utvecklas betydligt för att studien skall kunna ge några vetenskapligt acceptabla resultat.

Enligt min uppfattning vore detta en helt meningslös aktivitet, eftersom det inte räcker för en intressant empirisk vetenskaplig studie att den uppfyller de metodologiska kraven på en sådan. Det finns gott om metodologiskt korrekta men vetenskapligt fullständigt ointressanta beteendevetenskapliga studier 😊

De frågor som studeras måste därför också vara formulerade med en teoretisk förankring och en begreppsmässig precision som gör att de erhållna resultaten får inomvetenskapliga och/eller samhällliga konsekvenser. Men också här faller Turingtestet långt under den lägsta acceptabla nivån eftersom de centrala begreppen är både vaga och vetenskapligt ointressanta.

## Teoretiska/begreppsmässiga problem

Turing skriver i början på sin artikel att han inte vill diskutera om maskiner kan vara intelligenta, för det är ett alldeles för vagt begrepp. Det "parlor game" han föreslår är alltså enligt Turing själv inte ett test på om datorer kan ses som intelligenta. Han pratar också om huruvida datorer kan tänka, och många kommentatorer använder, precis som Turing "att tänka" och "att vara intelligent" som synonymer. Så oavsett vad Turing skriver har man i fortsättningen valt att se på Turingtestet som en metod för att avgöra om maskiner kan tänka eller kan ha intelligens.

Men varför i hela friden är detta intressanta frågor? Psykologer har sedan länge slutat att prata om "intelligens" som något monolitiskt enhetligt. Det som allmänheten kallar "intelligenstest" är för professionella psykologer, och inte bara forskare utan praktiker, ett batteri av tester av olika kognitiva förmågor (spatial, verbal, matematisk-logisk, aritmetisk m.m. <lägg in en korrekt lista>).

Läget är detsamma när det gäller att tänka. Det finns massor med distinktioner mellan olika slags tänkande. Typ 1 vs Typ 2, parallellt vs seriellt, operationellt vs figurativt (Piaget), visuellt vs verbalt osv. Och listan kan göras längre.

Min slutsats, och som jag tycker är uppenbart självklar (men en uppfattning som uppenbarligen inte delas av många, av allt att döma) är att det inte finns något som helst vetenskapligt intresse av att få veta svaret på om datorer kan "tänka" eller "är intelligenta". (Däremot verkar den delas av Turing själv, som i en sällan citerad del av sitt papper att *The original question "Can machines think" I believe to be too meaningless to deserve discussion*" (Turing, 1950<sup>1</sup>).

Däremot kan det finnas ett stort inte bara vetenskapligt utan också praktiskt-tillämpat intresse av ett kunna klargöra vilka kognitiva förmågor som datorer har eller kan ha och vilka de inte kan ha. Eller snarare, vilka de kan lika bra eller bättre än människor, och vilka förmågor där de presterar sämre än människor. Rent vetenskapligt skulle detta kunna vara en viktig pusselbit i försöken att öka vår förståelse för hur människans kognitiva förmågor fungerar rent procedurellt. Men än viktigare skulle en sådan kunskap leda till en mer balanserad diskussion om hur datorer kan användas i olika sammanhang i samhället, och inte minst i vilka sammanhang de på grund av sina svagheter absolut inte bör användas. Den diskussionen är en mycket viktig diskussion, som praktiskt taget inte existerar idag. Istället förs den offentliga diskussionen om de preteoretiska och därmed oprecisa och vaga termerna "tänkande" och "intelligens".

---

<sup>1</sup> Eftersom jag ej har originaltexten tillgänglig kan jag ej ange sidnummer. Citatet är taget från andra stycket i avsnittet **Contary Views on the Main Question**.

MANUS UNDER UTARBETANDE. CITERA EJ UTAN SKRIFTLIGT MEDIGIVANDE FRÅN FÖRFATTAREN.  
Men kommentera gärna!

Det teoretiskt oprecisa och metodologiskt ogenomtänkta Turingtestet är troligen en anledning till att den offentliga diskussionen om datorers kognitiva förmågor ägnar sig åt irrelevanta frågeställningar. Därför är det enligt min mening inte bara inomvetenskapligt olyckligt att så mycket diskussion om datorers kognitiva förmågor centrerats kring Turingtestet. Det innebär också att vi inte kan föra den viktiga, för att inte säga nödvändiga, diskussionen om datorteknikens påverkan på vårt framtida samhälle på ett konstruktivt sätt. Det är därför hög tid att flytta Turingtestet från dagens vetenskapliga och samhällliga diskussion om datorers förmågor och begränsningar till det ställe där det hör hemma. I historieböckerna.