

En kompromiss mellan kompatibilism och inkompatibilism

Ingmar Persson

Min utgångspunkt

Om två filosofiska positioner står mot varandra i en långvarig och segsliten debatt, är det en rimlig hypotes att de båda sitter på en del av sanningen. Av detta slag är min hypotes beträffande dispyten mellan kompatibilism och inkompatibilism med avseende på viljefrihet och ansvar (en hypotes som jag presenterar utförligast i Persson 2005, del V). Kompatibilism är tesen att viljefrihet och ansvar är förenliga med determinism och inkompatibilism tesen att de inte är det. Min hypotes skiljer mellan *ursprungligt* eller *yttersta* ansvar (*ultimate responsibility*) och *direkt* ansvar. Vi har direkt ansvar för avsiktliga handlingar som resulterar från avsikter som formats på basis av en begrundad av skäl som är fri i den bemärkelsen att den inte har störts av inre eller yttre tvång. Exempel på inre tvång är manier och fobier: personer som lider av kleptomani är inte fria att välja eller besluta sig för att låta bli att stjäla, och de som lider av agorafobi är inte fria att välja eller besluta sig för att korsa ett torg. Yttre tvång kan utgöras av pistolhot som medför att man väljer eller beslutar sig för att överlämna sina pengar till en vilt främmande individ, vilket man inte annars skulle ha gjort.

Genom avsikter som man bildar på detta fria eller otvungna sätt har man direkt kontroll över de avsiktliga handlingar som de manifesteras i. För att man ska vara ursprungligt eller ytterst ansvarig för en handling krävs inte bara denna direkta kontroll genom en avsikt utan direkt kontroll över alla de faktorer som bestämmer avsikten. Det krävs att det yttersta ursprunget till avsikten – och därmed den avsiktliga handlingen i vilken den implementeras – finns i agenten själv. Det krävs vad man skulle kunna kalla "egenmäktig" kontroll, en kontroll som kan grunda ansvar som är obetingat, inte betingat av några omständigheter som agenten inte råder över.

I en deterministisk värld är varelser som inte har existerat för evigt oförmögna att besitta egenmäktig kontroll. Kausalkedjor som utmynnar i avsikter som agenter med ett begränsat temporalt förflutet skaffar

sig måste i denna typ av värld ofrånkomligen löpa igenom dem till händelser som ägde rum före deras tillblivelse och som följaktligen är bortom deras kontroll. Om universum är indeterministiskt, behöver detta inte vara fallet: avsikter kan slumpmässigt ta form inom agenter. Problemet blir nu i stället att agenter inte kan ha direkt kontroll över dylika slumpmässiga händelser. Så oavsett om universum är deterministiskt eller indeterministiskt, kan agenter med våra temporala begränsningar inte vara utrustade med ursprungligt ansvar. Som kontrast finns ingen anledning att frånkänna oss direkt ansvar. De mekanismer som underminerar vårt ursprungliga ansvar verkar huvudsakligen i det fördolda, genom skeenden som vi till vardags inte är medvetna om. Frånvaron eller närvaron av de typer av tvång som kan beröva oss direkt ansvar är däremot vanligtvis observerbara i vardagen, och vi varseblir ofta deras frånvaro och känner oss därmed fria. Direkt ansvar är uppenbarligen förenligt med determinism; det är ju bara påverkan av en viss sorts orsaksfaktorer som utesluter sådant ansvar.

Ursprungligt ansvar för handlingar är emellertid det underlag som fordras för att vi ska kunna vara *förtjänta* av god eller dålig behandling på basis av handlingarna. Begreppet *förtjänst* involverar *rättvisa*: den belöning eller bestraffning som agenter förtjänar till följd av sina handlingar är en belöning eller bestraffning som det är *rättvist* att de erhåller till följd av dessa handlingar. Denna rättvisa består i en proportionalitet eller balans mellan värdet av deras handlingar och värdet för dem av belöningen eller bestraffningen. Men denna proportionalitet eller balans kan göra anspråk på att vara rättvis endast om handlingarna med sitt värde har sitt ursprung helt och hållet i agenterna. Likhet mellan värdet för agenterna av vissa belöningar eller bestraffningar och värdet av deras handlingar utgör rättvisa endast om handlingarna till fullo har sitt ursprung i agenterna. Om handlingarna i sista hand har sin källa i företeelser bortom agenterna, härrör deras värde ytterst därifrån och det blir orättvist att agenterna får erhålla belöningar eller bestraffningar som matchar deras handlingars värde.

Oavsett om världen är deterministisk eller indeterministisk, måste vi alltså revidera vårt ansvarsbegrepp så att det inte längre inbegriper förtjänst. Så långt är inkompatibilismen bärkraftig. Denna revision är emellertid förenlig med att vi bibehåller ett trimmat eller avskalat ansvarsbegrepp som är grundat på direkt kontroll, ett kompatibilistiskt ansvarsbegrepp som är förenligt med ett deterministiskt – men inte ett indeterministiskt – styre av vårt handlingsliv. Revisionen nödvändiggör

en utrensning av en hel del i vårt batteri av känslor som vrede, tacksamhet, stolthet, skam, beundran, förakt, skuld känsla och ånger. Alla dessa känslor involverar nämligen förtjänsttänkande. De som känner vrede mot någon tänker sig att vederbörande förtjänar någon form av vedergällning som en smäll på käften, medan de som känner tacksamhet mot någon tänker sig att vederbörande gjort sig förtjänt av någon lämplig belöning, och de som känner sig stolta eller skamsna tänker sig att de själva har gjort sig förtjänta av någon form av beröm respektive klander. Och så vidare.

Om man rensar bort så mycket av vår tanke- och känslors arsenal, kan man fråga sig om kvarstoden är tillräcklig för att vara värd namnet "ansvar"? Svaret är ja, för evolutionen har designat oss att tillskriva varandra förtjänst och att reagera med dessa förtjänstinfekterade känslor därför att det befrämjar vår överlevnad och reproduktion. Att vi reagerar med vrede och tillhörande aggressivt beteende gentemot dem som behandlar oss illa tenderar att avskräcka dem och vittnen till vårt argsinna gensvar från att behandla oss illa i framtiden, medan tacksamhet och ackompanjerande belöningsrespons gentemot dem som gynnat oss uppmuntrar till fortsatt välvilligt beteende gentemot oss. Så vår evolutionärt grundade praxis av ansvarshållande involverar dels ett tillbakablickande element som är förtjänstrelaterat – förtjänst är tillbakablickande därför att det är baserat på vad som har skett – dels ett framåtblickande element som består i de goda konsekvenserna för vårt vidkommande av att begagna det förtjänstrelaterade elementet. Förutom dessa goda konsekvenser hade våra förtjänststättigheter aldrig fortlevt, ifall de någonsin hade uppkommit. Så den föreslagna revisionen av ansvarsbegreppet slår vakt om en essentiell ingrediens i det; därför är det inte missvisande att fortsatt beteckna det med termen "ansvar".

Mitt avskalade ansvarsbegrepp ogiltigförklarar det tillbakablickande elementet i begreppet men bibehåller alltså det framåtblickande. Detta är ett möjligt förfarande eftersom vi nu har utvecklats intellektuellt till den grad att vi själva kan avgöra vilka ansvarsreaktioner som har bäst konsekvenser för oss, så vi är inte längre nödgade att förlita oss på de reaktioner som evolutionen har utrustat oss med. Vi har förmågan att inse att de evolutionärt betingade responserna kan vara alltför grovt tillyxade, exempelvis att vår ilska möjligtvis är alltför kraftig och det medföljande hämndbegäret alltför långtgående. Att vi inser att förtjänsttänkande och följaktligen de emotioner som inkorporerar det är irrationella innebär dock inte per automatik att de försvinner, för

de är så djupt evolutionärt inplanterade i oss att det är praktiskt taget omöjligt att utrota dem.

Som sagt, förtjänst är relaterat till rättvisa: det anses rättvist att vi får tillgodo vad vi förtjänar. Om förtjänstbegreppet är ohållbart, måste förstås gångbar rättvisa bestå i något annat. Jag har annorstädes argumenterat för att det består i *jämlikhet* (se Persson 2017, del II). Men oavsett hur rättvisa tolkas, kan det inte vara det *enda* moraliska värdet. Varelsers välbefinnande är ett annat moraliskt värde, vilket är förklaringen till varför ett rättvist tillstånd där varelsers välbefinnande är högt kan vara moraliskt bättre än ett likaledes rättvist tillstånd där det är lägre. De två moraliska värdena av rättvisa och grad av välbefinnande kan dock råka i konflikt med varandra, så att det ena måste vägas mot det andra, t. ex. mindre grad av rättvisa mot mer av välbefinnande. Som vi senare ska se, är det möjligt att dessa konflikter skärps om rättvisa omtolkas i termer av jämlikhet i stället för förtjänst.

Sapolskys position

Jag ämnar nu jämföra min position med den position Robert Sapolsky utvecklar i sin imponerande bok *Determined* (2023) eftersom det finns både viktiga likheter och skillnader mellan oss. Han skriver att en målsättning med hans bok är att ”visa hur ... determinism fungerar, att utforska hur biologin över vilken du inte hade någon kontroll interagerande med omgivningen över vilken du inte hade någon kontroll skapade dig” (s. 3, alla ospecificerade sidhänvisningar nedan är till nämnda bok och är översatta av mig). Denna programförklaring kan ge intryck av att han är omedveten om kvant-indeterminismen, men i realiteten betvivlar han bara att den ”bubblar upp tillräckligt för att förändra beteende”; om den mot förmodan förändrade vårt beteende, ”skulle allt den producerade vara slumpmässighet” (s. 238) som vi inte kontrollerar. I likhet med mig hävdar Sapolsky att både determinism och indeterminism utesluter moraliskt ansvar i en mening som inkluderar förtjänst (se t. ex. s. 11). Han anser att detta ställer honom inför ett ”moraliskt imperativ” att bli kvitt de (i föregående sektion exemplifierade) känsloreaktioner som förutsätter förtjänsttänkande – ett imperativ så svårt att åtlyda att han tillstår att han endast stundtals lyckas med det (s. 9).

Så långt gemensamma nämnare mellan Sapolsky och mig; nu till den väsentliga skillnaden: han erkänner inte något avskalat kompatibilistiskt ansvarsbegrepp vid sidan av det inkompatibilistiska. Han för-

nekar blankt att det finns viljefrihet – och därigenom implikationsvis den viljefrihet som jag menar kan manifesta sig i direkt ansvar – och deklarerar att det ”därför är fel att hålla folk moraliskt ansvariga för deras handlingar” (s. 11).

Att ett kategoriskt förnekande av moraliskt ansvar är ohållbart framgår emellertid tydligt av detta uttalande av honom: ”Om viljefrihet är en myt ... är våra handlingar bara det *amoraliska* utfallet av biologisk tur för vilket vi inte är ansvariga” (s. 246, min kursivering). Just det, om moraliskt ansvar inte finns, finns inte heller moraliska normer eller föreskrifter enligt vilka våra gärningar är moraliskt riktiga eller felaktiga – och alltså heller inte något ”moraliskt imperativ” att uttradera förtjänstfiltrerade emotioner. Liksom t. ex. knäreflexen (s. 314) eller epileptiska konvulsioner är alla våra handlingar varken det ena eller det andra. Sapolsky skriver mycket riktigt också att skiftet till den moderna synen på epileptiska anfall som fenomen för vilka vi inte är ansvariga ”förser oss med en fin modell för den mer globala uppgift som är central för boken” (s. 312), nämligen uppgiften att framställa våra avsiktliga handlingar som handlingar för vilka vi saknar allt ansvar.

Men, som han noterar, bara den som har ”viljefrihet, kunde ha valt att handla annorlunda” (s. 382). Enligt en välkänd slogan implicerar ”bör” ”kan”, varav det följer att om vi moraliskt bör utföra en handling, kan vi utföra den. Nu kan det uppenbarligen vara sant att vi bör utföra en handling, ehuru vi faktiskt handlar annorlunda. I så fall kan vi enligt Sapolsky inte (välja att) utföra handlingen, och följaktligen kan det inte vara sant att vi bör utföra den. Slutsats: Om vi saknar viljefrihet och ansvar, blir det meningslöst att tala om vad vi bör göra, ty detta kan inte vara något annat än vad vi faktiskt gör.

Detta gäller inte bara vad vi *moraliskt* bör göra. Sapolsky påstår: ”när folk tror mindre på viljefrihet, lägger de mindre avsiktlighet och ansträngning i sina handlingar, övervakar sina misstag mindre noggrant och satsar mindre på utfallet av en uppgift” (s. 248). Inte undra på, om de då betvivlar att de *kan* lägga mer avsiktlighet och ansträngning i sina handlingar (etc.). Detta äger tillämpning också på handlingar utanför den moraliska sfären, såväl kroppsliga handlingar som att sänka en golfputt som mentala aktiviteter i stil med att exekvera en matematisk uträkning. Skulle vi inte tro att vi kan göra bättre ifrån oss än vad vi faktiskt gör, blir det meningslöst att försöka förbättra sig.

Mot denna bakgrund, låt oss granska Sapolskys jämförelse av de konsekvenserna för moraliskt beteende av förnekelse av viljefrihet med dessa

konsekvenser av förnekelse av guds existens. Den senare förnekelsen kan ha moraliska konsekvenser av två olika skäl. Tanken kan för det första vara att om det inte finns någon gud som straffar överträdelser av moraliska bud, minskar människors benägenhet att hörsamma dessa bud. Det är synbarligen detta Sapolsky har i åtanke när han frågar: Kommer folk att bete sig mer omoraliskt när de konkluderar att "det inte finns Någon Allsmäktig som delar ut konsekvenserna" av sådant beteende? (s. 251–52). Hans slutsats är att i både med avseende på viljefrihet och guds existens har skepsis ringa effekt på moralisk livsföring (se t. ex. s. 264).

Vår tidigare konklusion att förnekelse av viljefrihet och ansvar omintetgör själva moralens existens visar emellertid att vad hans jämförelseobjekt borde vara är att förnekelse av guds existens kan ha moraliska konsekvenser av ett annat slag. Enligt en metaetisk teori är moraliska normer guds befallningar. Om detta äger riktighet, implicerar guds icke-existens att det inte existerar några moraliska normer och följaktligen att det blir omöjligt att handla såväl moraliskt riktigt som moraliskt felaktigt – alltså precis den konsekvens som jag har argumenterat inställer sig ifall man, som Sapolsky, förkastar viljefrihet av varje art som är nödvändig för moraliskt ansvar. Det säger sig självt att om det inte existerar moralnormer, finns det inget "behov av att uppfinna all-seende ögon i himmeln" (s. 252) med uppgift att övervaka åttlydnaden av dem.

Vi blir varse ytterligare problem med Sapolskys syn när vi tittar på hur han föreslår att brottslingar ska behandlas efter det att han har fränkännt dem ansvar för sina gärningar. Hans förslag drar en parallell med att sätta folk med farliga och smittsamma sjukdomar i karantän, så att de inte kan smitta andra (s. 349–51). Eftersom smittobärarna inte är ansvariga för sin farlighet för andra, bör begränsningen av deras rörelsefrihet vara så skonsam som möjligt, endast minimalt tillräcklig för att neutralisera deras farlighet. Detsamma gäller brottslingar: deras internering i fängelser bör påverka deras livskvalitet bara i så måtto att detta är resultatet av en inskränkning av deras rörelsefrihet som är nödvändig för att eliminera deras farlighet för andra. Därutöver får de inte bestraffas eller lida för sina ogärningar.

Sapolsky uppmärksammar en invändning mot denna modell (s. 353–54). Livskvaliteten försämras med stor sannolikhet för folk som place-ras i medicinsk karantän. Eftersom det som regel inte är deras egen skuld att de smittats, bör de i rättvisans namn kompenseras för den försämring reducerad rörelsefrihet innebär genom att erbjudas nya för-lustelsemöjligheter som ersättning för otillåtet socialt umgänge. När

storleken av denna kompensation kalibreras, bör man också beakta att redan det förhållandet att de blivit sjuka med största sannolikhet för med sig sänkt livskvalitet av ett slag som de genom sin isolering bidrar till att bespara sina medmänniskor ifrån. Denna sänkning berättigar ytterligare en höjning av deras kompensation.

Något liknande måste gälla för brottslingar som interneras, enligt Sapolsky. Eftersom de inte bär skuld för de brott de begått, bör de kompenseras för den inskränkning av deras handlingsfrihet som är oundgänglig för att skydda andra för deras farlighet, vilket är den enda form av sänkning av fångars välmående som är tillåten. Vidare, de faktorer som har förorsakat deras kriminalitet – t. ex. misshandel och sexuella övergrepp i barndomen – som de heller inte är ansvariga för, har med största sannolikhet gjort att välfärdsnivån i deras tidigare liv varit lägre än normalt. Detta motiverar en ökning av deras välfärd i fängelset.

Sapolsky kan ha rätt i att "världen skulle bli ett ännu mer rättvist ställe" (s. 316) om dessa reformer genomfördes, men problemet är att fängelsevistelsen då kan bli så behaglig att fångarna inte vill överge den. Frigivning eller hotet om frigivning skulle tendera att generera nya brott med syftet att uppnå ytterligare bekväm internering. Personer med låg levnadsstandard skulle bli frestade att slå in på brottets bana. (Motvarande problem skulle uppkomma om medicinsk karantän blev så tilltalande att folk lockades att frivilligt dra på sig smitta.)

Slutsatsen blir att syftet med fängelseförvaring inte kan vara blott och bart att skydda andra genom att avskilja internerna från dem. Förvaringen måste också innebära så pass låg livskvalitet att folk avskräcks från att begå brott. Jag håller med Sapolsky att detta kan innebära att man orättvist försämrar livslotten för kriminella individer som redan har haft sämre liv än flertalet. Detta kan man inte förneka om man förkastar det ursprungliga ansvar som bildar underlag för förtjänst och relaterad rättvisa. Denna orättvisa uppvägs emellertid av det positiva bidrag en sådan fängelseinstitution lämnar till tryggheten och välbefinnandet i samhället eller världen över lag. För att ta ett extremt exempel, antag att någon hade lyckats mörda Adolf Hitler redan i andra världskrigets begynnelse och därmed väsentligt hade förkortat det. Om vi ponerar att Hitlers tidigare liv hade varit huvudsakligen dåligt, skulle denna förlust av framtida levnadsår kunna innebära en ur jämlikhetsperspektiv orättvis försämring för honom, men denna orättvisa skulle mer än väl ha uppvägs av den avsevärda förbättring över lag som ett betydligt kortare världskrig hade utgjort.

För den stora vilseledda majoritet för vilken rättvisa består i att få vad man förtjänar, skulle mordet på Hitler kunna vara rättvist, ty han skulle ha förtjänat det i kraft av sin roll som andra världskrigets *primus motor*. Både Sapolskys och min position gör våld på förhärskande intuitioner genom att fördöma denna rättviseuppfattning och den tillfredsställelse som folk känner när de inbillar sig att brottslingar får vad de förtjänar som irrationella. Men, som jag här har försökt visa, far min kompromisslösning med sin acceptans av kompatibilistiskt ansvar grundat på direkt kontroll mer skonsamt fram med det ”sunda förnuftets” intuitioner än hans kategoriska avfärdande av *allt* ansvar.

Litteratur

- Persson, Ingmar. 2005. *The Retreat of Reason*. Oxford: Clarendon Press.
Persson, Ingmar. 2017. *Inclusive Ethics*. Oxford: Oxford University Press.
Sapolsky, Robert M. 2023. *Determined: A Science of Life Without Free Will*. London: Vintage.