

Om konsekventialism med grader av fel

Johan E. Gustafsson

I sin avhandling *Heavy Duty* presenterade Björn Eriksson (1994) en spännande utvidgning av konsekventialismen där handlingar kan vara fel i olika grad beroende på hur svårt det hade varit att utföra en bättre handling. Efter tänkvärda invändningar från Folke Tersman (1995) presenterade Eriksson (1996, 1997) en reviderad version av teorin i uppsatsen ”Utilitarism för syndare”. Denna reviderade version bemöttes sedan med en ytterligare runda invändningar av Tersman (1997). Efter denna kritik har, såvitt jag vet, inte teorin diskuterats mer. Mitt syfte i denna uppsats är att försöka råda bot på detta. Jag skall försöka visa hur Erikssons förslag med små förändringar kan räddas från Tersmans invändningar och en del andra problem.

Traditionell konsekventialism säger att en handling är rätt om, och endast om, dess konsekvenser är minst lika bra som de för varje alternativ handling och i fall den inte är rätt är den fel. De felaktiga handlingarna tycks dock rymma en stor variation som kan tänkas vara moraliskt relevant. Bland de felaktiga alternativen till en riktig handling kan värdena hos deras konsekvenser skilja sig mycket åt. Konsekvenserna för en av dem kan vara näst intill optimala medan en annans är katastrofala i jämförelse. Dessutom tycks en intuitivt relevant skillnad mellan olika felaktiga handlingar ligga i hur svårt det var att undvika dem—hur svårt det hade varit att i stället utföra en riktig handling. Eriksson vill utvidga konsekventialismen så att den blir kapabel att ta in denna typ av skillnader för att mer finkornigt gradera handlingars felaktighet.

En handlings grad av felaktighet beror enligt Eriksson på två faktorer nämligen, värdedifferens och relativ svårighet. Vi börjar med värdedifferens, som nog är det minst kontroversiella elementet i teorin. Betrakta en valsituation S_1 med tre alternativ vars konsekvenser har värden enligt följande:

S_1	
a_1	100
a_2	99
a_3	0

[s. 32] Här tycks det rimligt att säga att a_3 är mer fel än a_2 då förlusten i värde gentemot det optimala alternativet a_1 är större för a_3 än för a_2 . Vi inför följande notation för värdedifferens och felaktighet:

Vyx = värdet av konsekvenserna för y minus värdet av konsekvenserna för x .

Fx = x :s grad av felaktighet.

Teorin säger nu, givet att alla alternativ går att utföra utan svårighet, att

Fx = maximum av Vyx för alla alternativ y .¹

Det vill säga, graden av felaktighet för ett alternativ x är lika med den största möjliga differensen mellan något alternativ och x . Detta ger i exemplet ovan, givet att handlingarna alla går att utföra utan svårighet, att $Fa_3 = 100$ medan $Fa_2 = 1$. Detta tycks mig helt i linje med det utslag en konsekventialistisk teori kapabel hantera grader av felaktighet borde ge. Denna del av teorin tycks inte behöva revideras.

Det mest originella elementet i Erikssons teori torde vara idén att graden av felaktighet hos ett alternativ delvis beror på hur svårt det är att utföra ett bättre alternativ. Vi tänker oss att varje alternativ har en viss lätthetsgrad som kan uttryckas på en skala från 0, motsvarande omöjlig, till 1, motsvarande maximalt lätt, det vill säga, utförbar helt utan svårighet.² Eriksson (1996, s. 41) ger följande exempel där den upphöjda siffran representerar handlingens grad av lätthet:

S2	S3
a_1 100 ^{0,1}	a_1 100 ^{0,9}
a_2 10 ^{0,5}	a_2 10 ^{0,5}

Här är tanken att trots att värdedifferensen är densamma i situation S2 och S3 så är a_2 i S3 mer fel än a_2 i S2 eftersom det i S3 är lättare att utföra det bättre alternativet än det är i S2. Eriksson inför nu en funktion för den relativa svårigheten mellan två alternativ,

Syx = y :s grad av lätthet delat med x :s grad av lätthet.

[s. 33] Här väcks kanske en oro för division med noll om alternativ kan ha lätthetsgrad 0. En rimlig princip tycks dock vara att om en handling är

¹ Alternativt med matematisk notation: Om A är mängden av alla alternativ så är $Fx = \max\{Vyx : y \in A\}$.

² Här har jag tagit mig friheten att försöka förbättra terminologin. Det jag kallar grader av lätthet går hos Eriksson under namnet grader av svårighet. Det vill säga, svårighetsgrad 0 motsvarar omöjligt och svårighetsgrad 1 motsvarar något som kan utföras helt utan svårighet. Detta verkar dock bakvänt. Om till exempel en viss matteuppgift har en högre svårighetsgrad än en annan borde den ju vara svårare snarare än lättare.

ett alternativ så kan den utföras. Inte bara bör, utan även alternativskap, implicerar kan. Detta medför att alla alternativ måste ha en lätthetsgrad större än 0 och därmed att division med noll undviks.

Detta resonemang leder oss vidare till ytterligare ett skäl till att ta hänsyn till grader av lätthet. Handlingars lätthet är relevant för moralisk riktighet i alla rimliga moralteorier. De handlingar som är moraliskt relevanta är de som inte är så svåra att de är omöjliga att utföra. Så att låta moralisk riktighet och felaktighet bero på grader av lätthet är att mer finkornigt ta hänsyn till något som alla rimliga moralteorier tar hänsyn till. Ett sätt att tänka på det hela skulle kunna vara att alternativskap kommer i grader eller att en handling relevans som alternativ kommer i grader beroende på hur lätt den är att utföra. Detta tycks för mig som idéer väl värda att utforska.

Vi har nu allt vi behöver för att ställa upp Erikssons teori.

$$F_x = \text{maximum av } V_{yx} \text{ gånger } S_{yx} \text{ för alla alternativ } y.^3$$

Detta betyder att du räknar ut ett alternativ x :s grad av felaktighet enligt denna teori genom att först räkna ut produkten av V_{yx} och S_{yx} för alla alternativ y . Den största av dessa produkter är sedan x :s grad av felaktighet.

Eriksson (1996, s. 42) skriver att S_{yx} -funktionen utformats så att följande desiderata är uppfyllda: (a) Ju svårare det bättre alternativet y är jämfört med x , desto mindre fel tenderar x att vara. (b) Om x är svårare att utföra än det bättre alternativet y så ökar detta felaktighetsgraden hos x . (c) Om x är lättare att utföra än det bättre alternativet y så minskar detta felaktighetsgraden hos x . (d) Om x och y är lika svåra så inverkar inte deras respektive svårighet på x :s grad av felaktighet.⁴ Teorin ger utslaget att i S_2 så är $F_{a_2} = 18$ och i S_3 så är $F_{a_2} = 162$. Vi får alltså det önskade resultatet att a_2 är mer fel i S_3 än i S_2 .

Tersman (1997, s. 50) inleder sin uppsats med följande motexempel till Erikssons teori:

$$\begin{array}{l} S_4 \\ a_1 \quad 101^{0,1} \\ a_2 \quad 100^{0,001} \\ a_3 \quad 0^{0,899} \end{array}$$

³ Återigen, med matematisk notation: Om A är mängden av alla alternativ så är $F_x = \max\{V_{yx} \cdot S_{yx} : y \in A\}$.

⁴ I Erikssons beskrivning av sina desiderata jämför han inte, som jag skriver, med en bättre handling utan med den riktiga handlingen. Dock måste hans desiderata formuleras om i stil med ovan om de skall vara förenliga med hans slutgiltiga version av sin teori. Se Eriksson (1996, s. 44). Dessutom skriver Eriksson "lättare" i stället för "svårare" i (b), samt "svårare" i stället för "lättare" i (c). Detta måste vara ett misstag då det inte går ihop med formeln för S -funktionen. Felet är korrigerat i den engelska versionen av uppsatsen, se Eriksson (1997, s. 220).

[s. 34] Här ger Erikssons teori utslaget att $Fa_2 = 100$ och $Fa_3 \approx 11,2$. Det tycks orimligt att a_2 skulle vara nära på 9 gånger så fel som a_3 . Värdet av a_2 :s konsekvenser är nästan lika bra som de för den riktiga handlingen, a_1 , medan a_3 :s konsekvenser är mycket sämre. Att a_2 är mycket svårare än a_1 medan a_3 är relativt lätt borde inte göra a_2 mer fel än a_3 . Fall S4 visar att Erikssons teori bryter mot följande princip:

Konsekventialistisk dominans (KD)

Om två handlingar x och y är tillgängliga i samma situation och x har bättre konsekvenser än y så är x inte mer fel än y i situationen.

Om en handling har bättre konsekvenser än en annan så har den ju kommit närmare vad som enligt konsekventialismen är moralens mål, att konsekvenserna blir så bra som möjligt.⁵ Därmed borde handlingen med bättre konsekvenser inte bedömas som moraliskt värre. Ger man upp KD är det inte så mycket kvar av traditionell konsekventialism.

Fall av S4:s slag visar på det orimliga med att låta graden av felaktighet för ett visst alternativ bero på dess *egen* grad av lätthet. Mer relevant torde vara hur lätt det hade varit att få till det bättre utfallet. Ett kanske ännu tydligare utslag av detta finner vi i följande fall, där två felaktiga alternativ har lika bra konsekvenser men är olika svåra:

S5	
a_1	$100^{0,3}$
a_2	$10^{0,1}$
a_3	$10^{0,9}$

Erikssons teori ger att $Fa_2 = 270$ och $Fa_3 = 30$. Det tycks irrelevant för grad av felaktighet huruvida du lyckats utföra det svårare alternativet a_2 eller den lättare a_3 . Vad som spelar roll är att oavsett om du utförde a_2 eller a_3 så har du åstadkommit ett utfall värt 10 värdeenheter medan du hade kunnat åstadkomma ett utfall värt 90 mer med lätthetsgrad 0,3. Det vill säga, a_2 borde ha samma grad av felaktighet som a_3 .

Då a_2 och a_3 har lika bra konsekvenser men enligt Erikssons teori är olika mycket fel i S5, bryter teorin mot följande princip, som enligt Krister Bykvist (2002, s. 52) är en del av ”den konsekventialistiska andan”:
[s. 35]

Konsekventialistisk superveniens (KS)

Om två handlingar som är tillgängliga i samma situation har lika bra konsekvenser så har de samma deontiska status i situationen.

Ger man upp KS, ger man upp en tämligen grundläggande princip i konsekventialismen. Detta är kanske i sig inte så allvarligt, men så länge

⁵ Se Bentham (1970, s. 282) och Parfit (1984, s. 24).

vår ambition, likt Erikssons, är att åstadkomma en någorlunda konservativ utvidgning av konsekventialismen bör vi nog undvika detta.⁶

Är man ännu inte övertygad, notera att KS tycks svår att undvika om man accepterar KD samt följande princip som en konsekventialism med grader av fel borde uppfylla:

Konsekventialistisk kontinuitet (KK)

Små förändringar av värdet på en handlings konsekvenser i en situation ger inte, allt annat lika, stora utslag på de tillgängliga handlingarnas grader av felaktighet i situationen.

Här är tanken att om små förändringar av värdet på en handlings konsekvenser skulle ge stora utslag på handlingars grader av felaktighet trots att deras lätthet inte ändras, så är det väldigt svårt att tro att dessa grader mäter något av moralisk relevans för en konsekventialist. Små förändringar av de moraliskt relevanta faktorerna borde inte ge ett stort utslag på den moraliska bedömningen. KK torde åtminstone i detta sammanhang vara tämligen okontroversiell. Erikssons teori bryter inte heller mot KK.

För att se hur KD och KK stödjer KS, ta ett utslag som är oförenligt med KS där två handlingar med lika bra konsekvenser får olika grader av felaktighet. Till exempel, det utslag som Erikssons teori ger i S_5 , $Fa_2 = 270$ och $Fa_3 = 30$. Gör nu en godtyckligt liten förbättring av konsekvenserna till den av de två handlingarna som fick högst grad av felaktighet, i vårt exempel, a_2 . För att en teori med ett utslag likt det Erikssons gav inte skall bryta mot KD måste denna godtyckligt lilla förbättring ge ett så [s. 36] stort utslag på handlingarnas grader av felaktighet att Fa_2 inte längre är högre än Fa_3 . Det vill säga, en godtyckligt liten förändring i värde måste då ge ett stort utslag på grad av felaktighet, vilket bryter mot KK. Så om vi accepterar KD och KK borde vi även acceptera KS.⁷

⁶ Eriksson har meddelat mig att han gladeligen ger upp KS, då "en suboptimal handling som innebär att *man gör sig extra besvär* för att handla fel (det hade varit *lättare* att handla rätt) är mer fel än en axiologiskt ekvivalent handling som är lättare än att handla rätt." Svårt dock att skönja vad grunden för denna insikt är. Det tycks som att Eriksson blandar in fel sorts intuitioner, det vill säga, icke-konsekventialistiska intuitioner. Tanken skulle antingen kunna vara att svåra utföranden eller att man tar sig extra besvär skulle vara något som är i sig moraliskt problematiskt utöver eventuell påverkan på konsekvensernas värde. Eller så är den att det verkar särskilt tragiskt med en person som gör sig extra besvär för något som i slutändan ändå ger ett dåligt resultat. Ytterligare skulle det kunna tyckas ondskefullt att göra sig extra besvär för att, eller i syfte att, göra fel. Problemet är att inget av detta tycks kunna ges någon konsekventialistisk förankring.

⁷ För att KS strikt skall följa från KD och KK, måste dock KK preciseras en aning. Detta kommer att spela roll ifall en teori bryter mot KS men ändå implicerar att handlingar med lika bra konsekvenser alltid har nästan samma grad av felaktighet. Fler komplikationer uppstår också om vi tillåter att tillgängliga handlingar kan sakna en grad av felaktighet. Jag hoppas dock att ovanstående räcker för att förmedla den övergripande strukturen på argumentet.

Huruvida du utför en svår handling eller en lätt handling med lika bra konsekvenser torde vara moraliskt irrelevant för en konsekventialist. Detta utesluter dock inte att det ändå kan finnas rum för grader av lätthet även i en någorlunda konservativ utvidgning av konsekventialismen som håller fast vid KD, KK och KS. Enligt konsekventialismen beror som bekant ett alternativs deontiska status inte bara på värdet av dess konsekvenser. Värdet av andra handlingars konsekvenser är också relevanta givet att de inte är omöjliga att utföra i situationen. Det vill säga relevansen eller vikten av värdet av andra handlingars konsekvenser beror till viss mån på dessa handlingars lätthetsgrad. Detta öppnar upp för en utvidgning av konsekventialismen som mer finkornigt tar hänsyn till handlingars grad av lätthet utan att införa något alltför främmande element som inte i någon mån redan finns med i den traditionella tappningen.

Traditionell konsekventialism implicerar att ett alternativ är fel om, och endast om, det finns en handling med bättre konsekvenser som inte är omöjlig att utföra. Detta kan uttryckas som att ett alternativ x är fel om, och endast om, det finns en handling y sådan att Vyx gånger y :s grad av lätthet är större än noll. En naturlig utvidgning av detta vore att låta hur mycket fel ett alternativ är bero på hur mycket en sådan produkt av värdedifferens och det bättre alternativets grad av lätthet skiljer sig från noll. Jag föreslår att vi byter ut S -faktorn i Erikssons förslag mot det bättre alternativets grad av lätthet:

$$Fx = \text{maximum av } Vyx \text{ gånger } y\text{:s grad av lätthet för alla alternativ } y.^8$$

Med andra ord, för att räkna ut graden av felaktighet för ett alternativ x , räkna först ut, för alla alternativ y , produkten av Vyx och y :s grad av lätthet. Graden av felaktighet för x är lika med den största av dessa produkter. Detta reviderade förslag bryter inte mot KD. I S_4 ger det till exempel det rimligare utslaget att $Fa_2 = 0,1$ och $Fa_3 = 10,1$. Teorin bryter inte heller mot KS. I S_5 ger den till exempel utslaget $Fa_2 = Fa_3 = 27$.

[s. 37] Senare i sin uppsats ger Tersman (1997, s. 52) ett andra motexempel som vid första anblick tycks kräva mer drastiska revisioner av teorin.

S6		S7	
a_1	$1000^{0,5}$	a_1	$1000^{0,5}$
a_2	$0^{0,5}$	a_2	$0^{0,5}$
a_3	$500^{0,99}$		

Problemet, enligt Tersman, är här att Erikssons teori (likt min revision) implicerar att a_2 är lika fel i S_6 som i S_7 . Tersman menar att a_2 borde vara

⁸ Det vill säga, om Lx är x :s grad av lätthet och A är mängden av alla alternativ så är $Fx = \max\{Vyx \cdot Ly : y \in A\}$.

mer fel i S6 än i S7. Detta eftersom det i S6, utöver det riktiga alternativet, finns ytterligare ett alternativ, a_3 , som har avsevärt bättre konsekvenser än a_2 och som dessutom är extremt lätt att utföra.

Här tror jag dock att Tersman har fel. Notera först att alternativen här skall förstås som ömsesidigt uteslutande. Anta till exempel att möjligheterna stanna hemma, gå på bio och gå på fotboll är ömsesidigt uteslutande. Anta ytterligare att det är värre att välja att stanna hemma när man kunde ha gått på bio än att välja att stanna hemma när man kunde gått på fotboll. Jag har då svårt att se att det skulle kunna vara värre att stanna hemma om både bio och fotboll var alternativ än att välja att stanna hemma om bara bio finns som alternativ. Du hade ju fortfarande inte kunnat gå på både bio och fotboll. Förluster av missade möjligheter adderas helt enkelt inte ihop ifall dessa är sinsemellan uteslutande. Vare sig man har valt a_2 i S6 eller S7 så är a_1 den värsta missade möjligheten vad gäller missad värdedifferens i kombination med lätthet. Därmed borde inte a_2 vara mer fel i S6 än i S7.

Notera även att S6 är något av ett gränsfall. Hade värdet av a_3 :s konsekvenser i stället varit 506 eller högre så hade a_2 varit mer fel i S6 än i S7 enligt både Erikssons teori och min revision. Våra intuitioner om S6 och S7 är kanske därför inte så pålitliga.

Slutligen problematiserar Tersman (1997, s. 52) Erikssons (1996, s. 41) idé att de olika alternativens grad av lätthet i en situation alltid summerar till minst 1. Givet en naturlig tolkning av grader av lätthet tycks det tänkbart med situationer med två alternativ som båda är svårare än medelsvåra. Vilket torde resultera i en summa av lätthetsgrader lägre än 1. Hur som helst, en enkel lösning på problemet är helt enkelt att ge upp kravet på att lätthetsgraderna skall summera till minst 1. Inget viktigt i teorin hänger på detta krav.⁹ [s. 38]

Tack till Gustaf Arrhenius, Björn Eriksson, Nicolas Espinoza, Marc Fleurbaey, Jesper Jerkert och Martin Peterson för värdefulla kommentarer.

Litteratur

Bentham, Jeremy (1970), *An Introduction to the Principles of Morals and Legislation*, The Collected Works of Jeremy Bentham, Athlone, London.

Bykvist, Krister (2002), "Alternative Actions and the Spirit of Consequentialism", *Philosophical Studies* 107 (1): 45–68.

⁹ Även Erikssons egna exempel tycks bryta mot kravet. I situation A (här återgiven som S2) från Eriksson (1996, s. 41) finns det två handlingar, en medelsvår och en mycket svår. Om medelsvår lätthetsgrad motsvarar 0,5 så torde en mycket svår handling ha en ännu lägre lätthetsgrad varpå deras summa måste vara lägre än 1.

- Eriksson, Björn (1994), *Heavy Duty*, Almqvist & Wiksell, Stockholm.
- Eriksson, Björn (1996), "Utilitarism för syndare", *Filosofisk tidskrift* (4): 37–48.
- Eriksson, Björn (1997), "Utilitarianism for Sinners", *American Philosophical Quarterly* 34 (2): 213–228.
- Parfit, Derek (1984), *Reasons and Persons*, Clarendon Press, Oxford.
- Tersman, Folke (1995), "Recension av Björn Eriksson: *Heavy Duty*", *Filosofisk tidskrift* (3): 43–58.
- Tersman, Folke (1997), "Om Erikssons grader av synd", *Filosofisk tidskrift* (3): 50–53.