Den Gyllene Regeln och Substitutionsfunktioner

Daniel Rönnedal

Abstrakt
Enligt den s.k. gyllene regeln bör vi behandla andra så som vi själva vill bli behandlade. Det här är en princip som kan uttryckas och tolkas på många olika sätt. ”*Allt* ni vill att andra gör för er det skall ni också göra för dem” är en annan formulering. Denna sats antyder att regeln *kan* förstås på ett sådant sätt att den handlar om *alla* (typer av) handlingar. Men vad innebär det? I den här uppsatsen undersöker jag några olika preciseringar. Jag går igenom hur man kan förstå uttrycket ”alla handlingar” med hjälp av substitutionsfunktioner som ersätter handlingspredikat med välformade formler. Jag visar hur man med denna tolkning som utgångspunkt kan härleda en mängd normer som alla tycks följa ur den gyllene regeln. Enligt denna läsning är principen potentiellt mycket kraftfull och användbar. Tolkningen leder emellertid även till vissa tekniska problem som talar för att den är alltför stark. Jag visar därefter hur dessa problem kan lösas.

1. Introduktion

Enligt den s.k. gyllene regeln bör vi behandla andra så som vi själva vill bli behandlade. Den gyllene regeln är en av världens mest spridda, historiskt inflytelserika och accepterade normer. Alla världsreligioner, och många andra religioner, tycks innehålla någon variant av denna regel[1], och mängder av filosofer har också accepterat den i en eller annan form[2]. Det är emellertid inte alls uppenbart hur denna norm bäst uttrycks och tolkas. Det finns en mängd olika formuleringar av den gyllene regeln och varje uttryck i denna princip kan preciseras på många olika sätt.[3] Detta leder till bokstavligt talat tusentals olika möjliga tolkningar av denna välkända princip. Jag kommer i den här uppsatsen att undersöka *en* av de många frågor som uppstår då man

[1] Neusner och Chilton (red.) (2008).
[2] Se t.ex. Hobbes (1985), Kapitel XIV, s. 190, Pufendorf (1964), Bok 2, 3:13, Mill (1987), Kapitel 2, s. 28, och Gensler (1996), särskilt Kapitel 5.
[3] Rönnedal (2015) tar upp ett antal frågor som är relevanta då man försöker förstå den gyllene regeln. Se också Rönnedal (2016).

försöker förstå denna regel, nämligen huruvida den handlar om *alla* (typer av) handlingar eller bara *vissa* (typer av) handlingar. Uppsatsens huvudsyfte är att försöka förstå den logiska formen hos den gyllene regeln, inte att diskutera olika argument för eller emot denna princip. Personligen tror jag att det finns tolkningar av den gyllene regeln som är rimliga, även om det också finns många preciseringar som har problematiska konsekvenser. Oavsett om man är intresserad av att försvara eller kritiskt granska denna regel eller någon särskild precisering, bör man försöka förstå vad den innebär och fokusera på de bästa möjliga tolkningarna.[4]

Uppsatsen är indelad i sju avsnitt. Avsnitt 2 tar upp några interpretationer av den gyllene regeln och nämner ett antal slutsatser som tycks följa ur denna princip. Avsnitt 3 handlar om substitutionsfunktioner. I Avsnitt 4 visar jag hur teorin om substitutionsfunktioner kan användas för att härleda alla de satser som introduceras i Avsnitt 2 och som intuitivt följer ur GR. Avsnitt 5 innehåller ett problem för GR4, en viss tolkning av GR, som innebär att man kan härleda ett antal kontraintuitiva satser om man antar att man kan tillämpa vilka substitutionsfunktioner som helst på denna regel. Avsnitt 6 tar upp några möjliga lösningar på problemet ifråga. Jag argumenterar för att man kan undvika den aktuella svårigheten om man begränsar substitutionsfunktionernas värden till ”universella” predikat som inte innehåller några individkonstanter. Avsnitt 7 sammanfattar uppsatsen.

2. Tolkningar av den gyllene regeln

Enligt den kanske enklaste formuleringen av den gyllene regeln säger denna princip att om du vill att en individ x utför handling H mot dig, så bör du utföra H mot x. Och om vi generaliserar detta får vi följande resultat.

> **(GR)**. Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y.

---

[4] För mer historisk information om den gyllene regeln, se t.ex. Wattles (1996) och Gensler (2013), Kapitel 5. För en diskussion om den gyllene regelns förhållande till olika religioner, se Neusner och Chilton (red.) (2008). Filosofiska introduktioner till den gyllene regeln finner man bl.a. i Carson (2010), Kapitel 6, Carson (2013), Gensler (1996), särskilt Kapitel 5, och Gensler (2013). För mer information om den gyllene regeln se t.ex. Blackstone (1965), Bruton (2004), Cadoux (1912), Duxbury (2009), Gensler (1986), (2013), Gewirth (1978), Gould (1980), Hare (1963), Hertzler (1934), Hirst (1934), Hoche (1978), Huang (2005), Reinikainen (2005), Rönnedal (2015), Singer (1963), Wattles (1996) och Weiss (1941).

Detta ger oss dock inte i sig något svar på vad ”H” står för. GR formuleras ibland på följande sätt. *Allt* ni vill att andra gör för er bör ni också göra för dem. Denna formulering antyder att GR *kan* och kanske också *bör* tolkas på ett sådant sätt att den handlar om ”alla” (typer av) handlingar. Då skulle vi kunna uttrycka GR på följande sätt.

> **(GRH)**. Det gäller för alla handlingar H och individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y.

Men vad innebär det? Och vad menas med ”alla handlingar” i detta sammanhang? Inbegriper det alla handlingar över huvud taget eller alla handlingar av en viss typ? Låt oss nämna fyra av de intressantaste tolkningarna av detta uttryck.

Tolkning 1. ”alla handlingar” betyder alla handlingar som kan beskrivas med enkla handlingspredikat. Gör vi denna tolkning får vi följande variant av GR.

> **(GR1)**. Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för en handling som kan beskrivas med ett enkelt handlingspredikat.

Tolkar vi GR på detta sätt tycks regeln medföra bl.a. följande satser.

> (K1) Om du vill att din vän är ärlig mot dig, så bör du vara ärlig mot din vän.
>
> (K2) Om du vill att din partner är trogen mot dig, så bör du vara trogen mot din partner.
>
> (K3) Om du vill att denna främling håller sina löften till dig, så bör du hålla dina löften till denna främling.

Tolkning 2. ”alla handlingar” betyder alla handlingar som kan beskrivas med enkla handlingspredikat eller negationer av enkla handlingspredikat. Givet denna läsning, får vi följande variant av GR.

> **(GR2)**. Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för en handling som kan beskrivas med ett enkelt handlingspredikat eller en negation av ett enkelt handlingspredikat.

Tolkar vi GR på detta sätt tycks regeln medföra den s.k. silverregeln eller den ”negativa” formen av GR. Silverregeln kan uttryckas på följande sätt:

> **(SR)**. Det gäller för alla individer x och y att: Om x vill att y *inte* utför handling H mot x, så bör x *inte* utföra H mot y.

Dessutom tycks bl.a. följande satser vara konsekvenser av GR enligt denna tolkning.

> (K4) Om du vill att din granne *inte* stjäl från dig, så bör du *inte* stjäla från din granne.
>
> (K5) Om du vill att din arbetskamrat *inte* ljuger för dig, så bör du *inte* ljuga för din arbetskamrat.
>
> (K6) Om du vill att din ovän *inte* dödar dig, så bör du *inte* döda din ovän.

Tolkning 3. ”alla handlingar” betyder alla handlingar som kan beskrivas med enkla handlingspredikat, negationer av enkla handlingspredikat eller sådana handlingspredikat kopplade till ett villkor. Läser vi ”alla handlingar” på detta sätt, får vi följande variant av GR.

> **(GR3)**. Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för en handling som kan uttryckas med ett enkelt handlingspredikat, en negation av ett enkelt handlingspredikat eller ett handlingspredikat av sådant slag kopplat till ett villkor.

Tolkar vi GR på detta sätt tycks regeln medföra bl.a. följande satser.

> (K7) Om du vill att din bror hjälper dig om du behöver hjälp, så bör du hjälpa din bror om han behöver hjälp. (Om du vill att om du behöver hjälp så hjälper din bror dig, så bör det vara fallet att om din bror behöver hjälp så hjälper du honom.)
>
> (K8) Om du vill att din syster tackar dig om du hjälper henne, så bör du tacka din syster om hon hjälper dig. (Om du vill att om du hjälper din syster så tackar hon dig, så bör det vara fallet att om din syster hjälper dig, så tackar du henne.)

(K9) Om du vill att denna familjemedlem ber om ursäkt till dig om hon har svikit dig, så bör du be om ursäkt till denna familjemedlem om du har svikit henne. (Om du vill att om denna familjemedlem har svikit dig så ber hon om ursäkt till dig, så bör det vara fallet att om du har svikit denna familjemedlem så ber du om ursäkt till henne.)

Tolkning 4. GR talar enligt denna tolkning om alla handlingar över huvud taget (oavsett vilka handlingspredikat vi använder för att beskriva våra handlingar). Dvs. GR tolkas på följande sätt.

(GR4). Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för en handling som kan uttryckas med vilket handlingspredikat som helst.

Med denna tolkning tycks alla satserna (K1)–(K9) följa ur GR. Dessutom tycks t.ex. följande sats vara en konsekvens av GR4.

(K10) Om du vill att x behandlar alla dina barn med respekt, så bör du behandla alla x’s barn med respekt.

Låt mig nämna ytterligare några möjliga tolkningar. Alla varianter av GR ovan kan begränsas på så sätt att de endast handlar om vissa (typer av) handlingar, som t.ex. kan definieras genom en uppräkning. Den första varianten av GR skulle då kunna uttryckas på följande sätt.

(GR1′). Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för handlingen (i) …, (ii) … osv.

Där ”(i) …, (ii) … osv.” fylls i med en lista på de (typer av) handlingar man antar att GR uttalar sig om, t.ex. ”(i) att tala sanning, (ii) att hålla sina löften… osv.”. Silverregeln skulle kunna formuleras på följande sätt givet denna tolkning.

(SR′). Det gäller för alla individer x och y att: Om x vill att y inte utför handling H mot x, så bör x inte utföra H mot y, där H står för handlingen (i) …, (ii) …, (iii) … osv.

Där ”(i) …, (ii) …, (iii) … osv.” återigen fylls i med en lista på de (typer av) handlingar man antar att SR uttalar sig om, t.ex. ”(i) att ljuga, (ii) att stjäla, (iii) att döda… osv.”.

I de ovanstående tolkningarna av GR har vi explicit talat om ”handlingar”. Men GR kan också tolkas på sådant sätt att den inbegriper t.ex. attityder, känslor och förhållningssätt, förutom handlingar i en mer snäv mening. En konsekvens av GR tolkad på detta sätt skulle t.ex. kunna vara följande. Om du vill att din arbetskamrat *känner uppskattning* när du har hjälpt henne, så bör du *känna uppskattning* när din arbetskamrat har hjälpt dig.

Regeln kan också tolkas på ett sådant sätt att den även handlar om *hur* vi vill bli behandlade av andra och *hur* vi bör behandla andra. Det tycks ofta vara fallet att vi bryr oss om *hur* andra behandlar oss och inte bara *att* de utför en viss (typ av) handling. En konsekvens av GR tolkad på detta sätt skulle t.ex. kunna vara följande. Om du vill att x håller sitt löfte till dig *utan att klaga och vara allmänt otrevlig*, så bör du hålla ditt löfte till x *utan att klaga och vara allmänt otrevlig*.

Alla varianter av GR ovan ((GR1)–(GR4)) kan tolkas brett, så att de inkluderar olika (typer av) handlingar, attityder, känslor, förhållningssätt osv. Men de kan också begränsas till uttalanden om ”handlingar” i en mer snäv mening.

Tolkning 4 ger oss den starkaste formen av GR, tolkning 1 den svagaste. GR4 medför GR3, som medför GR2, som medför GR1. I en viss mening är alltså GR4 den intressantaste tolkningen. Inom vetenskapsteorin brukar man ofta betrakta en mer generell teori som bättre än en mindre generell teori (allt annat lika). Samma sak tycks gälla för normativa teorier. Men det finns också risk för att starka teorier blir alltför starka, så att vi kan härleda orimliga konsekvenser. Därför kan det vara rimligt att börja med att undersöka så starka teorier som möjligt och sedan i ljuset av eventuella kontraintuitiva slutsatser revidera dem. I en viss mening är det därför rimligast att börja med att undersöka tolkning 4.

Enligt tolkning 4 tycks alltså alla satserna (K1)–(K10) följa ur den gyllene regeln. Men *hur*? Det är givetvis enkelt att *påstå* att de följer. Men kan vi *bevisa* detta?

Kantianer hävdar ofta att det följer ur det kategoriska imperativet att vi bör hålla våra löften. Men det är ofta oklart exakt hur detta antas följa. De argument som åberopas är ofta informella. Och det är inte uppenbart om de kan omformuleras till deduktivt giltiga argument. Gäller detta även den typ av ”konsekvenser” av GR som vi har nämnt ovan?

Om vi använder tolkning 4, kan den logiska formen hos GR anges på följande sätt:

(FGRH) $\forall H \forall x \forall y(V_x Hyx \rightarrow OHxy)$.

Denna sats läses ”Det gäller för alla handlingar H och alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra handling H mot y”. Vi kvantifierar här explicit över handlingar. En alternativ formulering är den följande:

(FGRS) $\forall x \forall y(V_x Hyx \rightarrow OHxy)$,

där H tolkas schematiskt, som en variabel som kan bytas ut mot vilket predikat som helst. I praktiken säger detta schema samma sak som $\forall H \forall x \forall y(V_x Hyx \rightarrow OHxy)$.

Använder vi någon av dessa symboliseringar följer alla satserna (K1)–(K10), eller – mer precist – adekvata formaliseringar av dessa satser. Gör vi det uppstår emellertid en del problem, bl.a. samma problem som om vi antar att vi kan ersätta handlingspredikatet H i den gyllene regeln med vilken formel som helst (se Avsnitt 5 nedan). Dessutom krävs det, åtminstone för den första symboliseringen, att vi utvecklar en högre ordningens logik. Det är intressant att se om vi kan undvika det.

Vi skall i den här uppsatsen istället visa hur alla satserna (K1)–(K10) kan härledas med hjälp av substitutionsfunktioner från handlingspredikatet H till (godtyckliga) formler. Närmare bestämt kommer vi att begränsa oss till substitution av formler som inte innehåller några individkonstanter för att undvika det problem med GR4 som diskuteras i Avsnitt 5. Det här leder till en tolkning av den gyllene regeln, nämligen GR5, som är något svagare än GR4. GR4 medför denna tolkning, men inte tvärtom. För att förstå detta lösningsförslag måste vi först veta lite mer om substitutionsfunktioner. Vi börjar gå igenom den nödvändiga teorin i nästa avsnitt. Den ”teori” för dessa funktioner vi skall använda har utvecklats av S. C. Kleene (1971) och Gerhard Schurz (1997). Vi följer Schurz framställning nästan ordagrant, men vi kommer att göra några kompletteringar.

3. Substitutionsfunktioner

För att kunna uttrycka oss exakt kommer vi att utveckla ett formellt språk, som vi sedan använder för att formalisera den gyllene regeln och härleda satserna (K1)–(K10).

Syntax

Konventioner

σ och π denoterar substitutionsfunktioner för icke-logiska predikat, den första ersätter predikat med formler (”komplexa” predikat), den andra ersätter predikat med predikat. Notera att σ tar den formel som förekommer omedelbart till höger som argument. $\sigma Fu_1 \ldots u_n$ står t.ex. för $\sigma(Fu_1 \ldots u_n)$, och $\sigma A \rightarrow B = \sigma(A) \rightarrow B$, som inte är identisk med $\sigma(A \rightarrow B) = \sigma(A) \rightarrow \sigma(B) = \sigma A \rightarrow \sigma B$. Om vi vill uttrycka att σ refererar till ett predikat, kan vi skriva σF, $\sigma(F)$, $(\sigma F)[u_{1\text{-}n}/x_{1\text{-}n}]$ eller $(\sigma(F))[u_{1\text{-}n}/x_{1\text{-}n}]$.

$\mathcal{L}$, $\mathcal{V}$, $\mathcal{C}$, $\mathcal{T}$, $\mathcal{R}$… denoterar mängden av alla formler, variabler, konstanter, termer, respektive relationer. A, B, C… betecknar godtyckliga formler.

Schurz använder x, y, z inte endast som individvariabler i objektspråket utan också som metavariabler i metaspråket, som varierar över alla individvariabler; vi skall också låta s, t, u, v generellt variera över termer. Detsamma gäller predikatvariabler F, G, H… och satsvariabler p, q, r… Schurz använder även objektspråkets satslogiska konnektiv och kvantifikatorer i metaspråket. ”:=” står för identitet per definition. **N** står för mängden av alla naturliga tal. Vi använder samma konventioner i den här uppsatsen.

Vokabulär

Schurz (1997, s. 34) introducerar endast en mängd individvariabler utan en extra mängd individkonstanter. Han klarar sig därmed utan distinktionen mellan öppna och slutna formler. Givet de vanliga distinktionerna mellan fria och bundna variabler, spelar de fria individvariablerna i Schurz system samma roll som individkonstanter i språk med en distinktion mellan individvariabler och individkonstanter. Det grundläggande språk Schurz beskriver innehåller följande vokabulär.

(1) en uppräkningsbart oändlig mängd $\mathcal{V}$ av individvariabler u, v, ... , x, y, z (möjligtvis med index).

(2) För varje $n \geq 0$, en uppräkningsbart oändlig mängd $\mathcal{R}^n$ av n-ställiga predikat F, G, H … (möjligtvis med index); i synnerhet gäller det att $\mathcal{P} := \mathcal{R}^0$ står för mängden av satsvariabler p, q, … (noll-ställiga predikat). $\mathcal{R} := \cup_{n \geq 0} \mathcal{R}^n$ denoterar mängden av alla predikat. …

(3) De logiska symbolerna (konnektiven och operatorerna) $\neg$ (negation), $\vee$ (disjunktion), $\forall$ (universell kvantifikator), $\square$ (aletisk nödvändighets operator), O (deontisk plikt operator), och parenteser), och (.

Vi skall emellertid utvidga detta alfabet på två sätt. Vi introducerar

(4) en uppräkningsbart oändlig mängd $\mathcal{C}$ av individkonstanter a, b, c ... (möjligtvis med index); och

(5) en (logisk) satsoperator V (en viljeoperator).

Anledningen till detta är att vi vill kunna referera till specifika individer med hjälp av en mängd konstanter och att vi vill kunna uttrycka att en eller flera individer *vill* någonting. Variablerna och konstanterna kallas tillsammans för ”termer”. Mängden av alla termer betecknas $\mathcal{T}$.

Definition 1 (Icke primitiva symboler). Symbolerna $\rightarrow$ (materiell implikation), $\wedge$ (konjunktion), $\leftrightarrow$ (materiell ekvivalens), $\top$ (Verum), $\bot$ (Falsum), $\exists$ (existens kvantifikator), $\Diamond$ (aletisk möjlighets operator) och P (deontisk tillåtelse operator) definieras på vanligt sätt. $(A \rightarrow B) := (\neg A \vee B)$, $(A \wedge B) := \neg(\neg A \vee \neg B)$, $(A \leftrightarrow B) := ((A \rightarrow B) \wedge (B \rightarrow A))$, $\top := (p \vee \neg p)$, $\bot := \neg\top$, $\exists xA := \neg\forall x\neg A$, $\Diamond A := \neg\square\neg A$, $PA := \neg O\neg A$. Notera att varje individvariabel som är bunden räknas som en logisk symbol, medan fria individvariabler och predikat räknas som icke-logiska symboler. ■

Språk

Schurz (1997, ss. 34–35) identifierar ett språk $\mathcal{L}$ med mängden av alla formler som genereras av följande regler:

(i) $F \in \mathcal{R}^n, x_1, \ldots, x_n \in \mathcal{V} \Rightarrow Fx_1 \ldots x_n \in \mathcal{L}$ (de atomära formlerna);
(ii) $A, B \in \mathcal{L} \Rightarrow \neg A, (A \vee B), \square A, OA \in \mathcal{L}$;
(iii) $A \in \mathcal{L}, x \in \mathcal{V} \Rightarrow \forall xA \in \mathcal{L}$; och
(iv) Ingenting annat är en formel.

Vi skall emellertid modifiera dessa regler på följande sätt. Vi byter ut (i) mot (i′), (iv) mot (iv′), och lägger till (v).

(i′) $F \in \mathcal{R}^n, t_1, \ldots, t_n \in \mathcal{T} \Rightarrow Ft_1 \ldots t_n \in \mathcal{L}$ (de atomära formlerna);
(iv′) $A \in \mathcal{L}, t \in \mathcal{T} \Rightarrow V_tA \in \mathcal{L}$;
(v) Ingenting annat är en formel.

Kommentar 2. De satslogiska konnektiven läses på vanligt vis. □A läses ”Det är nödvändigt att A”, OA ”Det bör vara fallet att A” (eller ”Det är obligatoriskt att A”), och V_tA ”t vill att A”.

Parenteser utelämnas i regel om ingen mångtydighet uppstår. Predikatens ställighet bestäms av kontexten. I ”Fx” är F ett ett-ställigt predikat och i ”Fxy” ett två-ställigt predikat, etc. $\forall x_{1\text{-}n}A$ är en förkortning av $\forall x_1 \ldots \forall x_nA$; och likadant med $\exists x_{1\text{-}n}A$.

Om **Φ** är en term, formel eller mängd formler i $\mathcal{L}$, så betecknar $\mathcal{V}(\mathbf{\Phi})$, $\mathcal{C}(\mathbf{\Phi})$, $\mathcal{T}(\mathbf{\Phi})$, $\mathcal{R}(\mathbf{\Phi})$, $\mathcal{R}^n(\mathbf{\Phi})$, $\mathcal{P}(\mathbf{\Phi})$ mängden variabler, konstanter, termer, predikat, n-ställiga predikat, respektive satsvariabler som förekommer i **Φ**. $\mathcal{L}(\mathbf{\Phi})$ denoterar **Φ**’s språk, dvs. den mängd formler som kan konstrueras från $\mathcal{R}(\mathbf{\Phi}) \cup \mathcal{T}(\mathbf{\Phi})$ plus de logiska konnektiven och operatorerna. ■

Definitioner av viktiga begrepp

Begreppen fri och bunden variabel, alfabetisk variant etc. förklaras som vanligt. Om x är en variabel som förekommer i en formel A, så är det viktigt att skilja mellan variabeln x själv och en eller flera förekomster av x i A. En förekomst av x i A är en viss del av formeln A, en del som har den syntaktiska formen ”x”. Satsen $A := \forall xFx \wedge Gx$ innehåller t.ex. tre förekomster av variabeln x: den längst till vänster, omedelbart till höger om kvantifikatorn, är bunden, den i mitten är bunden, och den längst till höger är fri. På samma sätt måste vi skilja mellan en delformel B av A och förekomsterna av B i A. Satsen $A := B \wedge (B \rightarrow C)$ innehåller två förekomster av delformeln B.

Nu kan vi kalla en förekomst av variabel x i en formel A fri om och endast om (omm) den inte förekommer inom räckvidden för en förekomst av $\forall x$ i A (annars är den bunden). Vi kallar variabeln x själv fri i A omm den har åtminstone en fri förekomst i A; annars är x bunden i A. $\mathcal{V}_f(A)$ denoterar mängden av fria variabler i A och $\mathcal{V}_b(A)$ mängden av bundna variabler i A. Uppenbarligen gäller det att $\mathcal{V}_f(A) \cap \mathcal{V}_b(A) = \varnothing$ och att $\mathcal{V}_f(A) \cup \mathcal{V}_b(A) = \mathcal{V}(A)$. Begreppet variabel-substitution kan nu förklaras på vanligt sätt. Vi säger att y är fri för x i A (där $x, y \in \mathcal{V}$; $A \in \mathcal{L}$) omm x inte förekommer i A inom räckvidden för en kvantifikator som binder y. Givet att y är fri för x i A, så är A[y/x] – resultatet av den korrekta substitutionen av y för x i A – den formel som resulterar från A genom att ersätta varje fri förekomst av x med y. Om y är fri för x i A, så säger vi också att A[y/x] är definierad, annars är den odefinierad. På liknande sätt gäller det att $A[y_1/x_1, \ldots, y_n/x_n]$ denoterar resultatet av den korrekta simultana substitutionen av y_i för x_i (för alla $1 \leq i \leq n$; där x_i är parvis distinkta, men y_i inte behöver vara det). $A[y_{1\text{-}n}/x_{1\text{-}n}]$ är en förkortning av $A[y_1/x_1, \ldots, y_n/x_n]$. Notera att operationen ”[y/x]” tar hela formeln ”A” i ett uttryck av formen A[y/x] som argument. Schurz använder ibland parenteser för att belysa detta: t.ex. $\exists xA[y/x]$ läses $(\exists xA)[y/x]$, vilket är skilt från $\exists x(A[y/x])$ om A innehåller x fri (då är $\exists x(A[y/x]) \neq \exists xA$, medan $(\exists xA)[y/x] = \exists xA$) (Schurz (1997, s. 36)). Vi skall här göra likadant. Låt oss definiera detta lite mera exakt.

Definition 3 (fria och bundna variabler, öppna och slutna formler).

(1) En förekomst av en variabel v i A är fri i A omm den inte förekommer inom räckvidden för ett kvantifikator-uttryck av typen $\forall v$ eller $\exists v$.

Rekursiv definition

(i) Om A är atomär, så är varje förekomst av varje v i $\mathcal{V}$ fri i A.

(ii) Om $A = \neg B$, $\Box A$ eller OA, så är en v-förekomst fri (bunden) i A omm den är fri (bunden) i B.

(iii) Om $A = B \vee C$, så är en v-förekomst i B (C) fri (bunden) i A omm den är fri (bunden) i B (C).

(iv) Om $A = V_xB$, så är en v-förekomst i B fri (bunden) i A omm den är fri (bunden) i B; x-förekomsten omedelbart till höger om V i A är (alltid) fri i A. Om c är en konstant och $A = V_cB$, så är en v-förekomst fri (bunden) i A omm den är fri (bunden) i B.

(v) Om $A = \forall xB$, så är en v-förekomst fri i A omm v inte är identisk med x och v är fri i B, annars är v bunden i A.

(2) En variabel är fri i A omm den har åtminstone en fri förekomst i A. $\mathcal{V}_f(A)$ = mängden av variabler som är fria i A.

(3) En formel i $\mathcal{L}$ som inte har några fria variabler kallas en sluten formel eller sats. Annars är det en öppen formel eller sats. ■

Vi vill att A[t/x] skall vara den formel som uppstår från en substitution av t för alla fria förekomster av x i A. Vi vill alltså att A[t/x] skall vara en formel som säger samma sak om t som A säger om x.

Exempel 4. Om $A = Fx \rightarrow Hyx$, så är $A[t/x] = Ft \rightarrow Hyt$. Om $A = Fxy \rightarrow \forall yHyx$, så är $A[t/x] = Fty \rightarrow \forall yHyt$. Om $A = Fxy \rightarrow \forall xHyx$, så är $A[t/x] = Fty \rightarrow \forall xHyx$. ■

För att uppnå detta måste vi dock undvika att t blir bunden av någon kvantifikator $\forall t$ eller $\exists t$ i A efter substitution. Vi måste undvika ”inkorrekta” ”substitutioner” av följande slag.

Exempel 5. Om $A = Fxy \rightarrow \forall yHyx$, så är $A[y/x] = Fyy \rightarrow \forall yHy\underline{y}$. Om $A = \forall x\forall y(Fxy \rightarrow Gxy)$, så är $A[y/x] = \forall x\forall y(F\underline{y}y \rightarrow G\underline{y}y)$. Dessa ”substitutioner” är inkorrekta. Alla understrukna variabelförekomster är bundna av en kvantifikator i A. ■

Låt oss nu gå igenom några fler definitioner.

Definition 6 (Fri substitution).

(1) t är fri (att substitueras) för x i A omm ingen fri förekomst av x i A ligger inom räckvidden för en kvantifikator som binder t.

Rekursiv definition

(i) Om A är atomär, så är t fri för varje x i $\mathcal{V}$ i A.

(ii) t är fri för x i $\neg A$, $\Box A$, OA och V_sA om t är fri för x i A.

(iii) t är fri för x i $A \vee B$ om t är fri för x i A och i B.

(iii) t är fri för x i $\forall zA$ om t inte är z och t är fri för x i A.

(2) Om t är fri för x i A, så denoterar $(A)[t/x]$ resultatet av den simultana substitutionen av t för alla fria x-förekomster i A. I annat fall är $(A)[t/x]$ odefinierad. ■

Kommentar 7. Notera att $(A)[t/x]$ är en metalingvistisk notation. Parenteser kring A används för att undvika mångtydighet. Men vi skall också ofta utelämna dem. $A[t/x]$ opererar då på hela A, som vi tidigare påpekat. ■

Alla modala predikatlogiska system som Schurz beskriver är slutna under två substitutionsoperationer: substitution av fria individvariabler, och substitution av predikat (Schurz (1997, s. 45)). Dvs. i alla system Schurz beskriver gäller det att om A är ett teorem i logiken L, så är $A[y_{1\text{-}n}/x_{1\text{-}n}]$ ett teorem i L, givet att $A[y_{1\text{-}n}/x_{1\text{-}n}]$ är definierad. Schurz utvecklar en ny substitutionsoperation för predikat. En komplikation är att predikat inte endast kan ersättas av ordinära (atomära) predikat utan också av komplexa predikat, dvs. komplexa formler.

Notationen för uniform substitution av predikat har utvecklats av Kleene för icke-modal predikatlogik (Kleene (1971, ss. 155–162)), och utvidgas till modal predikatlogik av Schurz (1997, s. 45). Betrakta ett n-ställigt predikat F som följs av vissa parvis distinkta variabler $x_1, \ldots, x_n$; $x_1, \ldots, x_n$ i $Fx_1 \ldots x_n$ kallas för ”namn form” variabler (Kleene (1971, s. 156)). Uniform substitution av formeln B för $Fx_1 \ldots x_n$ innebär att varje förekomst av $Fu_1 \ldots u_n$ (där u_i är variabler i $\mathcal{V}$, inte nödvändigtvis distinkta) ersätts av den korresponderande B-substitutions-instansen $B[u_{1\text{-}n}/x_{1\text{-}n}]$ givet att vissa restriktioner är uppfyllda som förhindrar sammanblandning av variabler. Kom ihåg att $B[u_{1\text{-}n}/x_{1\text{-}n}]$ står för den sats som är resultatet av att samtidigt ersätta varje fri förekomst av x_1 mot u_1, varje fri förekomst av x_2 mot u_2 osv. De fria variabler i B som inte är namn form variabler, kallas anonyma variabler i B (Kleene (1971, s. 156f)). Notera att det är möjligt att B inte innehåller några fria förekomster av x_i ($1 \leq i \leq n$). I detta fall ersätts varje $Fu_1 \ldots u_n$ av samma formel B. Eftersom valet av namn form variabler är godtyckligt, så använder Schurz en oändlig mängd parvis distinkta namn form variabler $x_1, \ldots, x_i, \ldots$.

Definition 8 (Substitution av predikat).
En substitutionsfunktion för predikat är en funktion $\sigma: \mathcal{R} \rightarrow \mathcal{L}$. En formel B är fri för predikat F i $\mathcal{R}^n$ i en formel A omm det för varje $Fu_1 \ldots u_n$ i A gäller att:

(i) $B[u_{1\text{-}n}/x_{1\text{-}n}]$ är definierad, och

(ii) $Fu_1 \ldots u_n$ inte förekommer inom räckvidden för en kvantifikator som binder en förekomst av en anonym variabel i B (Kleene (1971, s. 156)). En formel A är fri för σ om det för varje F i $\mathcal{R}(A)$ gäller att σF är fri för F i A. Om A är fri för σ, så säger vi också att σA är definierad, annars är σA odefinierad. Om A är fri för σ, så denoterar σA resultatet av den simultana substitutionen av varje förekomst av $Fu_1 \ldots u_n$ med $(\sigma F)[u_{1\text{-}n}/x_{1\text{-}n}]$ i A för varje F i $\mathcal{R}^n(A)$ och n i **N**. (Schurz (1997, s. 46)) ■

Definition 9 (Rekursiv definition av ”A är fri för σ” och ”σA”).

(1) Om $A = Fu_1 \ldots u_n$ är atomär, så är A fri för σ omm $u_1, \ldots, u_n$ är fria för $x_1, \ldots, x_n$ i σF, resp.; och om detta är fallet, så är $\sigma A = (\sigma F)[u_{1\text{-}n}/x_{1\text{-}n}]$.

(2) Då $A = \neg B$, $B \vee C$, $\Box B$, OB, eller V_tB gäller följande: Om $A = \neg B$, $\Box B$, OB eller V_tB, så är A fri för σ omm B är fri för σ; i det kvarvarande fallet om B och C är fria för σ; och givet att A är fri för σ, så är $\sigma A = \neg\sigma B$, $(\sigma B \vee \sigma C)$, $\Box\sigma B$, $O\sigma B$, respektive $V_t\sigma B$.

(3) $A = \forall zB$. Då är A fri för σ om B är fri för σ och för varje G i $\mathcal{R}^n(B)$ (och varje n i **N**), z inte är en anonym variabel i σG; och om detta är fallet, så är $\sigma A = \forall z\sigma B$. (Schurz (1997, s. 46)) ■

Kommentar 10. Om $Fu_1 \ldots u_n$ är atomär, så är $\sigma(Fu_1 \ldots u_n) = (\sigma(F))[u_{1\text{-}n}/x_{1\text{-}n}]$, vilket kan förenklas till följande uttryck $\sigma(Fu_1 \ldots u_n) = (\sigma F)[u_{1\text{-}n}/x_{1\text{-}n}]$ eller $\sigma F[u_{1\text{-}n}/x_{1\text{-}n}]$.

Notera att genom att ersätta definierade symboler med primitiva begrepp så är det enkelt att se att induktiva villkor av typ (2) också gäller för $\wedge$, $\rightarrow$, $\leftrightarrow$, $\Diamond$ och P, och villkor av typ (3) för $\exists$.

Substitutionsfunktioner för propositioner är ett särskilt fall av substitutionsfunktioner för predikat begränsade till noll-ställiga predikat. ■

Några exempel

Diskussionen har hittills varit ganska abstrakt. Vi skall nu ta upp några exempel på konkreta substitutionsfunktioner för att belysa innebörden i de olika begreppen. Alla exempel, utom det första, är lånade från Schurz.

Låt F och H vara ett-ställiga predikat, och låt $\pi(F) = \pi F = G$, $\pi(H) = \pi H = H$ och $A := \forall x(Fx \rightarrow Hx)$. Här är π en substitutionsfunktion som tar oss från predikat till predikat. Då är $\pi(A) = \pi A = \pi(\forall x(Fx \rightarrow Hx)) = \pi\forall x(Fx \rightarrow Hx) = \forall x(Gx \rightarrow Hx)$. Vi kan stegvis komma fram till detta på följande sätt. $\pi(\forall x(Fx \rightarrow Hx)) = \forall x\pi((Fx \rightarrow Hx)) = \forall x(\pi(Fx) \rightarrow \pi(Hx)) = \forall x(\pi(F)x \rightarrow \pi(H)x) = \forall x(Gx \rightarrow Hx)$.

Låt F, och G vara ett-ställiga predikat, H, I, J, och K två-ställiga predikat; vidare, låt $\sigma(F) = \sigma F = Hx_1z$, $\sigma(G) = \sigma G = Izx_1$, $\sigma(J) = \sigma J = (\exists zHx_1z \rightarrow Kx_1x_2)$, och $A := \forall y((Fy \wedge Gy) \rightarrow \Box\exists uJyu)$. Notera att i $\sigma F = Hx_1z$ är x_1 en namn form variabel och z en anonym variabel, i $\sigma G = Izx_1$ är x_1 en namn form variabel och z en anonym variabel, och i $\sigma J = (\exists zHx_1z \rightarrow Kx_1x_2)$ är x_1 och x_2 namn form variabler och z en anonym variabel. Då är A fri för σ, och $\sigma(A) = \sigma A = \forall y((Hyz \wedge Izy) \rightarrow \Box\exists u(\exists zHyz \rightarrow Kyu))$. Men $B := Fu \rightarrow Jzu$ är inte fri för σ eftersom $\sigma J[z/x_1, u/x_2] = (\exists zHx_1z \rightarrow Kx_1x_2)[z/x_1, u/x_2]$ är odefinierad. $(\exists zHx_1z \rightarrow Kx_1x_2)[z/x_1, u/x_2] = (\exists zH\underline{z}z \rightarrow Kzu)$, och i $(\exists zH\underline{z}z \rightarrow Kzu)$, blir den understrukna förekomsten av z bunden av $\exists z$. $C := \forall z(Gu \wedge \neg Fu)$ är inte heller fri för σ eftersom den anonyma variabeln z i σG och σF blir bunden av $\forall z$ i σC (= $\forall z(I\underline{z}u \wedge \neg Hu\underline{z})$). Schurz nämner även ett exempel på en "degenererad" substitution. Om F och G är tvåställiga predikat och $\sigma F = Hx_1$, $\sigma G = p$, så är $\sigma\forall x\exists y(Fxy \rightarrow Gyy) = \forall xEy(Hx \rightarrow p)$ (som är ekvivalent med $\exists xHx \rightarrow p$).

Genom att använda den rekursiva definitionen av substitutionsfunktioner kan man stegvis komma fram till att σA ovan är $\forall y((Hyz \wedge Izy) \rightarrow \Box\exists u(\exists zHyz \rightarrow Kyu))$ på följande sätt.

$$
\begin{aligned}
&\sigma(\forall y((Fy \wedge Gy) \rightarrow \Box\exists uJyu)) = \\
&\forall y\sigma(((Fy \wedge Gy) \rightarrow \Box\exists uJyu)) = \\
&\forall y(\sigma((Fy \wedge Gy)) \rightarrow \sigma(\Box\exists uJyu)) = \\
&\forall y((\sigma(Fy) \wedge \sigma(Gy)) \rightarrow \Box\sigma(\exists uJyu)) = \\
&\forall y((\sigma(Fy) \wedge \sigma(Gy)) \rightarrow \Box\exists u\sigma(Jyu)) = \\
&\forall y(((\sigma(F))[y/x_1] \wedge (\sigma(G))[y/x_1]) \rightarrow \Box\exists u(\sigma(J))[y/x_1,u/x_2]) = \\
&\forall y(((Hx_1z)[y/x_1] \wedge (Izx_1)[y/x_1]) \rightarrow \Box\exists u(\exists zHx_1z \rightarrow Kx_1x_2)[y/x_1,u/x_2]) \\
&= \\
&\forall y((Hyz \wedge Izy) \rightarrow \Box\exists u(\exists zHyz \rightarrow Kyu))
\end{aligned}
$$

Om man utelämnar den yttersta parentesen runt en sats när ingen mångtydighet uppstår blir $\forall y\sigma(((Fy \wedge Gy) \rightarrow \Box\exists uJyu))$ istället $\forall y\sigma((Fy \wedge Gy) \rightarrow \Box\exists uJyu)$ och $\forall y(\sigma((Fy \wedge Gy)) \rightarrow \sigma(\Box\exists uJyu))$ istället $\forall y(\sigma(Fy \wedge Gy) \rightarrow \sigma(\Box\exists uJyu))$. $\forall y(((\sigma(F))[y/x_1] \wedge (\sigma(G))[y/x_1]) \rightarrow \Box\exists u(\sigma(J))[y/x_1,u/x_2])$ kan förenklas till $\forall y(((\sigma F)[y/x_1] \wedge (\sigma G)[y/x_1]) \rightarrow \Box\exists u(\sigma J)[y/x_1,u/x_2])$, vilken i sin tur kan förenklas till $\forall y((\sigma F[y/x_1] \wedge \sigma G[y/x_1]) \rightarrow \Box\exists u\sigma J[y/x_1,u/x_2])$.

Vi är nu redo att visa att satserna (K1)–(K10) i Avsnitt 2 är härledbara från GR4.

4. Lösningar av hur GR medför satserna (K1)–(K10) i Avsnitt 2

I Avsnitt 2 nämnde vi några satser ((K1)–(K10)) som alla tycks följa från den gyllene regeln, åtminstone om vi tolkar denna regel som (GR4). I det här avsnittet skall vi visa hur man kan förstå detta. Antag att den gyllene regeln symboliseras på följande sätt.

(FGR) $\forall x\forall y(V_xHyx \rightarrow OHxy)$

$\forall x\forall y(V_xHyx \rightarrow OHxy)$ läses ”Det gäller för alla x och y att: Om x vill att y utför H mot x, så bör x utföra H mot y”. Om vi antar att predikatet H i (FGR) kan ersättas med vilken formel som helst, kan vi härleda adekvata symboliseringar av alla satserna (K1)–(K10) i Avsnitt 2. Detta antyder att följande formella framställning av (GR4) är rimlig. Den gyllene regeln tolkad som (GR4) medför (FGR) och alla satser som kan fås från (FGR) med hjälp av en substitutionsfunktion σ, givet att (FGR) är fri för σ.

Låt oss i detalj gå igenom hur man kan härleda (K1), (K5), (K7) och (K10) om man tolkar den gyllene regeln på detta sätt.[5]

Härledning av (K1)
(K1) Om du vill att din vän är ärlig mot dig, så bör du vara ärlig mot din vän.

Låt Rxy stå för x är ärlig mot y och låt d vara en singulär term som refererar till dig och v en singulär term som refererar till din vän. Låt $\sigma(H) = Rx_1x_2$. x_1 och x_2 är namn form variabler, det finns ingen anonym variabel i uttrycket, och (FGR) är fri för σ. Då kan (K1) härledas med hjälp av nedanstående steg.

$$\sigma(\forall x\forall y(V_xHyx \rightarrow OHxy)) = \forall x\sigma(\forall y(V_xHyx \rightarrow OHxy)) =$$
$$\forall x\forall y\sigma((V_xHyx \rightarrow OHxy)) = \forall x\forall y(\sigma(V_xHyx) \rightarrow \sigma(OHxy)) =$$
$$\forall x\forall y(V_x\sigma(Hyx) \rightarrow O\sigma(Hxy)) =$$
$$\forall x\forall y(V_x(\sigma(H))[y/x_1, x/x_2] \rightarrow O(\sigma(H))[x/x_1, y/x_2]) =$$
$$\forall x\forall y(V_x(Rx_1x_2)[y/x_1, x/x_2] \rightarrow O(Rx_1x_2)[x/x_1, y/x_2]) =$$
$$\forall x\forall y(V_xRyx \rightarrow ORxy)$$

Slutsatsen i denna ”deduktion” kan läsas på följande sätt. ”Det gäller för alla x och y att: Om x vill att y är ärlig mot x, så bör x vara ärlig mot y”. Från detta följer:

$$V_dRvd \rightarrow OHdv$$

om vi instanserar x med d och y med v. $V_dRvd \rightarrow OHdv$ läses: ”Om du vill att din vän är ärlig mot dig, så bör du vara ärlig mot din vän” = (K1). V.S.B.

Härledning av (K5)
(K5) Om du vill att din arbetskamrat inte ljuger för dig, så bör du inte ljuga för din arbetskamrat.

Låt Lxy stå för x ljuger för y och låt d vara en singulär term som refererar till dig och a en singulär term som refererar till din arbetskamrat. Låt $\sigma(H) = \neg Lx_1x_2$. x_1 och x_2 är namn form variabler, det finns ingen anonym variabel i uttrycket, och (FGR) är fri för σ. (K5) kan nu härledas på följande sätt.

[5] I en strikt mening visar vi hur ett antal *symboliseringar* av (K1), (K5), (K7) och (K10) följer från (FGR) med hjälp av substitution av predikatet H. Men givet att dessa symboliseringar är ”korrekta”, kan vi dra slutsatsen att dessa satser själva följer ur GR.

$\sigma(\forall x\forall y(V_xHyx \to OHxy)) = \forall x\sigma(\forall y(V_xHyx \to OHxy)) =$
$\forall x\forall y\sigma((V_xHyx \to OHxy)) = \forall x\forall y(\sigma(V_xHyx) \to \sigma(OHxy)) =$
$\forall x\forall y(V_x\sigma(Hyx) \to O\sigma(Hxy)) =$
$\forall x\forall y(V_x(\sigma(H))[y/x_1, x/x_2] \to O(\sigma(H))[x/x_1, y/x_2]) =$
$\forall x\forall y(V_x(\neg Lx_1x_2)[y/x_1, x/x_2] \to O(\neg Lx_1x_2)[x/x_1, y/x_2]) =$
$\forall x\forall y(V_x\neg Lyx \to O\neg Lxy)$

Den sista satsen läses: ”Det gäller för alla x och y att: Om x vill att y inte ljuger för x, så bör det vara fallet att x inte ljuger för y”. Det följer att:

$V_d\neg Lad \to O\neg Lda$

om vi instanserar x med d och y med a. $V_d\neg Lad \to O\neg Lda$ läses: ”Om du vill att din arbetskamrat inte ljuger för dig, så bör du inte ljuga för din arbetskamrat” = (K5). V.S.B.

Härledning av (K7)
(K7) Om du vill att din bror hjälper dig om du behöver hjälp, så bör du hjälpa din bror om han behöver hjälp.
(Om du vill att om du behöver hjälp så hjälper din bror dig, så bör det vara fallet att om din bror behöver hjälp så hjälper du honom.)

Låt Bx stå för x behöver hjälp, Hxy för x hjälper y och låt d vara en singulär term som refererar till dig och b en singulär term som refererar till din bror. Låt $\sigma(H) = Bx_2 \to Hx_1x_2$. x_1 och x_2 är namn form variabler, det finns ingen anonym variabel i uttrycket, och (FGR) är fri för σ. Följande deduktion visar hur (K7) kan härledas ur (FGR) med hjälp av σ.

$\sigma(\forall x\forall y(V_xHyx \to OHxy)) = \forall x\sigma(\forall y(V_xHyx \to OHxy)) =$
$\forall x\forall y\sigma((V_xHyx \to OHxy)) = \forall x\forall y(\sigma(V_xHyx) \to \sigma(OHxy)) =$
$\forall x\forall y(V_x\sigma(Hyx) \to O\sigma(Hxy)) =$
$\forall x\forall y(V_x(\sigma(H))[y/x_1, x/x_2] \to O(\sigma(H))[x/x_1, y/x_2]) =$
$\forall x\forall y(V_x(Bx_2 \to Hx_1x_2)[y/x_1, x/x_2] \to O(Bx_2 \to Hx_1x_2)[x/x_1, y/x_2]) =$
$\forall x\forall y(V_x(Bx \to Hyx) \to O(By \to Hxy))$

Dvs. ”Det gäller för alla x och y att: Om x vill att y hjälper x om x behöver hjälp, så bör x hjälpa y om y behöver hjälp”. Det följer att:

$V_d(Bd \rightarrow Hbd) \rightarrow O(Bb \rightarrow Hdb)$

om vi instanserar x med d och y med b. $V_d(Bd \rightarrow Hbd) \rightarrow O(Bb \rightarrow Hdb)$ är en rimlig formalisering av (K7) = ”Om du vill att om du behöver hjälp så hjälper din bror dig, så bör det vara fallet att om din bror behöver hjälp så hjälper du honom”.

(K7) kan alltså fås genom substitution från den gyllene regeln: Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y. V.S.B.

Härledning av (K10)

(K10) Om du vill att x behandlar alla dina barn med respekt, så bör du behandla alla x’s barn med respekt.

Låt Bxy stå för x är ett barn till y, Rxy för x behandlar y med respekt, och låt d vara en konstant som refererar till dig. Låt $\sigma(H) = \forall z(Bzx_2 \rightarrow Rx_1z)$. $\sigma(H)$ läses ”Det gäller för alla z att om z är ett barn till x_2, så behandlar x_1 z med respekt”. Uttrycket saknar anonyma variabler, x_1 och x_2 är namn form variabler, och (FGR) är fri för σ. Följande härledning visar hur (K10) kan härledas ur (FGR) med hjälp av σ.

$$\sigma(\forall x\forall y(V_xHyx \rightarrow OHxy)) = \forall x\sigma(\forall y(V_xHyx \rightarrow OHxy)) =$$
$$\forall x\forall y\sigma((V_xHyx \rightarrow OHxy)) = \forall x\forall y(\sigma(V_xHyx) \rightarrow \sigma(OHxy)) =$$
$$\forall x\forall y(V_x\sigma(Hyx) \rightarrow O\sigma(Hxy)) =$$
$$\forall x\forall y(V_x(\sigma(H))[y/x_1, x/x_2] \rightarrow O(\sigma(H))[x/x_1, y/x_2]) =$$
$$\forall x\forall y(V_x(\forall z(Bzx_2 \rightarrow Rx_1z))[y/x_1, x/x_2] \rightarrow$$
$$O(\forall z(Bzx_2 \rightarrow Rx_1z))[x/x_1, y/x_2]) =$$
$$\forall x\forall y(V_x\forall z(Bzx \rightarrow Ryz) \rightarrow O\forall z(Bzy \rightarrow Rxz))$$

Den sista satsen i denna härledning läses: ”Det gäller för alla x och y att om x vill att det gäller för alla z att om z är ett barn till x så behandlar y z med respekt, så bör det vara fallet att det gäller för alla z att om z är ett barn till y så behandlar x z med respekt. Från denna sats kan vi sluta oss till följande formel om vi instanserar x med d och y med x.

$V_d\forall z(Bzd \rightarrow Rxz) \rightarrow O\forall z(Bzx \rightarrow Rdz)$

Denna formel läses: ”Om du vill att det gäller för alla z att om z är ett barn till dig så behandlar x z med respekt, så bör det vara fallet att det gäller för alla z att om z är ett barn till x så behandlar du z med respekt”. Och denna sats är synonym med (K10). V.S.B.

(K10) är en ganska komplicerad konsekvens av GR4 och visar att inte endast enkla kategoriska plikter följer från denna princip, utan också t.ex. olika sorters villkorliga normer. Det här gör GR4 till en mycket kraftfull moralisk maxim, som potentiellt kan användas i härledningen av många andra normer.

Notera också att alla konsekvenser av GR4 i sin tur kan användas i olika härledningar av olika moraliska påståenden. Betrakta t.ex. följande argument.

1. Den gyllene regeln.
2. Du vill att x skall behandla dina barn med respekt.
3. (Det är ett historiskt faktum att) Lisa är ett barn till x.

Alltså

4. Du bör behandla Lisa med respekt.

Detta argument är giltigt om vi antar att den gyllene regeln tolkas på det sätt vi har föreslagit ovan, vi använder en vanlig möjlig värld-semantik för vårt aletiskt-deontiska system, alla historiska fakta är historiskt nödvändiga, vi kvantifierar över alla möjliga individer, och alla deontiskt tillgängliga världar är aletiskt tillgängliga. Detta innebär inte nödvändigtvis att GR *faktiskt* är sann. Men det innebär att denna regel tillsammans med övriga premisser medför slutsatsen. Och alla andra premisser är plausibla.

Vi kan visa detta på följande sätt. 1 medför (i) $V_d\forall z(Bzd \rightarrow Rxz) \rightarrow O\forall z(Bzx \rightarrow Rdz)$. 2, $V_d\forall z(Bzd \rightarrow Rxz)$, tillsammans med (i) medför (ii) $O\forall z(Bzx \rightarrow Rdx)$. (ii) tillsammans med 3, $\Box Blx$, medför 4, ORdl. Att (ii) och 3 medför 4 kan visas på följande sätt. (Vi antar en vanlig möjlig värld-semantik i följande argument.) Antag att (1) $O\forall z(Bzx \rightarrow Rdx)$ och (2) $\Box Blx$ är sanna i den möjliga världen w_0, och att (3) ORdl är falsk i w_0. Då gäller det att (4) det finns en möjlig värld w_1 sådan att w_1 är deontiskt tillgänglig från w_0 [Från 3], och (5) Rdl är falsk i w_1 [Från 3]. Alltså, (6) $\forall z(Bzx \rightarrow Rdz)$ är sann i w_1 [Från 1 och 4]. (7) $Blx \rightarrow Rdl$ är sann i w_1 [Från 6]. (8) w_1 är aletiskt tillgänglig från w_0 [Från antagandet att alla deontiskt tillgängliga världar också är aletiskt tillgängliga]. Alltså, (9) Blx är sann i w_1 [Från 2 och

8]. Det följer att, (10) Rdl är sann i w_1 [Från 7 och 9], och att (11) $Rdl \wedge \neg Rdl$ är sann i w_1 [Från 5 och 10]. Men detta är absurt.

5. Problem för GR4: substitutionsargumentet

Den gyllene regeln är alltså en potentiellt mycket användbar moralisk princip. Vi skall emellertid nu undersöka ett argument som talar för att GR4 är en alltför stark tolkning av GR. Enligt den aktuella preciseringen av GR är alla satser som kan fås från (FGR) med hjälp av en substitutionsfunktion σ, givet att (FGR) är fri för σ, konsekvenser av den gyllene regeln. Men betrakta nu följande exempel.

Låt Fxy stå för x är ärlig mot y, d referera till dig och a till någon godtycklig, konkret person (t.ex. din partner). Antag att $\sigma(H) = (x_1 = a \wedge Fx_1x_2) \vee (\neg x_1 = a \wedge \neg Fx_1x_2)$. x_1 och x_2 är namn form variabler, och a är en individkonstant. Uttrycket innehåller inga anonyma variabler. Då är följande argument giltigt, om vi gör vissa antaganden som förefaller vara mycket plausibla (se nedan).

1. V_dFad	[Antag]
2. $V_d\sigma(H)[a/x_1, d/x_2]$	[Från 1, ”def av” $\sigma(H)[a/x_1, d/x_2]$ etc.]
3. $O\sigma(H)[d/x_1, a/x_2]$	[Från 2 med GR4]
4. $O\neg Fda$	[Från 3, ”def av” $\sigma(H)[d/x_1, a/x_2]$ etc.]

Här följer en informell läsning av detta argument.

Du vill att a skall vara ärlig mot dig.
Alltså vill du att a skall utföra σH mot dig.
Om du vill att a skall utföra σH mot dig, så bör du utföra σH mot a.
Alltså bör du utföra σH mot a.
Alltså bör du inte vara ärlig mot a.

Men denna slutsats är kontraintuitiv. Låt oss kalla detta argument för ”substitutionsargumentet”. Från det faktum att du vill att a skall vara ärlig mot dig och att GR4 är sann, följer det då att du *inte* bör vara ärlig mot a. Vi kan härleda liknande resultat för alla typer av handlingar. Om du vill att a är trogen mot dig, så bör du *inte* vara trogen mot a. Om du vill att a håller sina löften till dig, så bör du *inte* hålla dina löften till a osv. Detta är moraliskt orimligt.

Och inte nog med det. Vi kan dessutom härleda en direkt motsägelse om GR4 är sann och vi antar att det inte finns några genuina moraliska dilemman. För vi kan också härleda OFda på följande sätt. (SL innebär att steget följer med hjälp av vanlig satslogik.)

5. V_dFad	[Antag]
6. $V_dFad \rightarrow OFda$	[Från GR4]
7. $OFda$	[5, 6, SL]

Och från detta är det lätt att härleda en kontradiktion.

8. $OFda \wedge O\neg Fda$	[4, 7, SL]
9. $\neg(OFda \wedge O\neg Fda)$	[OD]
10. $(OFda \wedge O\neg Fda) \wedge \neg(OFda \wedge O\neg Fda)$	[8, 9, SL]

Men 10 är en motsägelse. De enda regler, förutom GR4, vi behöver anta i steg 5–10 är (OD), $\neg(OA \wedge O\neg A)$, som utesluter förekomsten av explicita moraliska dilemman, och olika grundläggande satslogiska regler. Och alla dessa principer tycks vara rimliga. Om argumentet är giltigt och alla övriga premisser är sanna, måste vi förkasta GR4. Det här argumentet är därför djupt problematiskt för denna tolkning av den gyllene regeln. Kan en anhängare av GR4 möjligtvis attackera steg 2, 3, eller 4 i argumentet ovan? Vi skall nu se hur dessa steg kan försvaras.

Steg 2 $V_dFad \Rightarrow V_d\sigma(H)[a/x_1, d/x_2]$. För att visa steg 2, måste vi visa att $V_dFad \Rightarrow V_d\sigma(H)[a/x_1, d/x_2]$. Notera att $\sigma H = (x_1 = a \wedge Fx_1x_2) \vee (\neg x_1 = a \wedge \neg Fx_1x_2)$. Alltså är $\sigma(H)[a/x_1, d/x_2] = (a = a \wedge Fad) \vee (\neg a = a \wedge \neg Fad)$. Det följer att $V_d\sigma(H)[a/x_1, d/x_2] = V_d((a = a \wedge Fad) \vee (\neg a = a \wedge \neg Fad))$. Så, för att visa att $V_dFad \Rightarrow V_d\sigma(H)[a/x_1, d/x_2]$, måste vi visa att $V_dFad \Rightarrow V_d((a = a \wedge Fad) \vee (\neg a = a \wedge \neg Fad))$. Detta följer om vi antar att V_xB följer ur V_xA om A och B är logiskt ekvivalenta (kalla denna regel (VE)). För Fad är logiskt ekvivalent med $(a = a \wedge Fad) \vee (\neg a = a \wedge \neg Fad)$. Regeln att V_xB följer ur V_xA om A och B är logiskt ekvivalenta är knappast sann för alla personer i alla situationer. Men det är rimligt att anta att alla personer som är fullständigt rationella satisfierar denna princip. Och om vi antar att den gyllene regeln gäller också för fullständigt rationella personer, så har vi fortfarande ett problem. Om vi antar att V fungerar som en normal modal operator, kan vi även bevisa att V_dFad medför $V_d\sigma(H)[a/x_1, d/x_2]$ på följande sätt.

$V_dFad \Rightarrow V_d\sigma(H)[a/x_1, d/x_2] = V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad))$

$$
\begin{array}{c}
(1)\ V_dFad, 0 \\
(2)\ \neg V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad)), 0 \\
(3)\ 0s1 \\
(4)\ \neg((a = a \land Fad) \lor (\neg a = a \land \neg Fad)), 1 \\
(5)\ \neg(a = a \land Fad), 1 \\
(6)\ \neg(\neg a = a \land \neg Fad), 1 \\
(7)\ Fad, 1 \\
\swarrow \qquad\qquad \searrow \\
(8)\ \neg a = a, 1 \qquad (9)\ \neg Fad, 1 \\
(10)\ * \qquad\qquad (11)\ *
\end{array}
$$

Steg 3 $V_d\sigma(H)[a/x_1, d/x_2] \Rightarrow O\sigma(H)[d/x_1, a/x_2]$. Vi har redan visat att $V_d\sigma(H)[a/x_1, d/x_2] = V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad))$. $\sigma H = (x_1 = a \land Fx_1x_2) \lor (\neg x_1 = a \land \neg Fx_1x_2)$. Alltså, $\sigma(H)[d/x_1, a/x_2] = (d = a \land Fda) \lor (\neg d = a \land \neg Fda)$. Således gäller det att $O\sigma(H)[d/x_1, a/x_2] = O((d = a \land Fda) \lor (\neg d = a \land \neg Fda))$. För att visa att $V_d\sigma(H)[a/x_1, d/x_2] \Rightarrow O\sigma(H)[d/x_1, a/x_2]$, måste vi alltså visa att $O((d = a \land Fda) \lor (\neg d = a \land \neg Fda))$ följer från $V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad))$ med hjälp av GR. Vi visar först att $\forall x\forall y(V_x((y = a \land Fyx) \lor (\neg y = a \land \neg Fyx)) \rightarrow O((x = a \land Fxy) \lor (\neg x = a \land \neg Fxy)))$ följer från (FGR) om vi tillämpar σ på denna sats. Denna sats medför i sin tur $V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad)) \rightarrow O((d = a \land Fda) \lor (\neg d = a \land \neg Fda))$ med hjälp av vanlig predikatlogik. Från detta följer omedelbart vårt resultat. Här är vårt bevis.

$\sigma H = (x_1 = a \land Fx_1x_2) \lor (\neg x_1 = a \land \neg Fx_1x_2)$.
$\sigma(\forall x\forall y(VxHyx \rightarrow OHxy)) = \forall x\sigma(\forall y(VxHyx \rightarrow OHxy)) =$
$\forall x\forall y\sigma((VxHyx \rightarrow OHxy)) = \forall x\forall y(\sigma(VxHyx) \rightarrow \sigma(OHxy)) =$
$\forall x\forall y(V_x\sigma(Hyx) \rightarrow O\sigma(Hxy)) =$
$\forall x\forall y(V_x(\sigma(H))[y/x_1,x/x_2] \rightarrow O(\sigma(H))[x/x_1,y/x_2]) =$
$\forall x\forall y(V_x((x_1 = a \land Fx_1x_2) \lor (\neg x_1 = a \land \neg Fx_1x_2))[y/x_1,x/x_2] \rightarrow$
$O((x_1 = a \land Fx_1x_2) \lor (\neg x_1 = a \land \neg Fx_1x_2))[x/x_1,y/x_2]) =$
$\forall x\forall y(V_x((y = a \land Fyx) \lor (\neg y = a \land \neg Fyx)) \rightarrow O((x = a \land Fxy) \lor (\neg x =$
$a \land \neg Fxy)))$.

$V_d((a = a \land Fad) \lor (\neg a = a \land \neg Fad)) \rightarrow O((d = a \land Fda) \lor (\neg d = a \land \neg Fda))$ är en instans av den sista satsen i härledningen ovan (låt x = d, och y = a).

Steg 4 $O\sigma(H)[d/x_1, a/x_2] \Rightarrow O\neg Fda$. $\sigma H = (x_1 = a \wedge Fx_1x_2) \vee (\neg x_1 = a \wedge \neg Fx_1x_2)$. Det innebär att $\sigma(H)[d/x_1, a/x_2] = (d = a \wedge Fda) \vee (\neg d = a \wedge \neg Fda)$ och att $O\sigma(H)[d/x_1, a/x_2] = O((d = a \wedge Fda) \vee (\neg d = a \wedge \neg Fda))$. För att visa $O\sigma(H)[d/x_1, a/x_2] \Rightarrow O\neg Fda$, måste vi alltså visa att $O((d = a \wedge Fda) \vee (\neg d = a \wedge \neg Fda))$ medför $O\neg Fda$. För att visa detta lägger vi till premissen att du inte är identisk med a, dvs. $\neg d = a$, och antar att all identitet är nödvändig. All identitet är nödvändig om vi antar att individkonstanterna tolkas som rigida designatorer och refererar direkt till individer. Här följer ett bevis av steg 4.

$O\sigma(H)[d/x_1, a/x_2], \neg d = a \Rightarrow O\neg Fda$

$$
\begin{array}{c}
(1)\ O((d = a \wedge Fda) \vee (\neg d = a \wedge \neg Fda)),\ 0 \\
(2)\ \neg d = a,\ 0 \\
(3)\ \neg O\neg Fda,\ 0 \\
(4)\ P\neg\neg Fda,\ 0 \\
(5)\ 0s1 \\
(6)\ \neg\neg Fda,\ 1 \\
(7)\ (d = a \wedge Fda) \vee (\neg d = a \wedge \neg Fda),\ 1 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
(8)\ d = a \wedge Fda,\ 1 & (9)\ \neg d = a \wedge \neg Fda,\ 1 \\
(10)\ d = a,\ 1 & (11)\ \neg d = a,\ 1 \\
(12)\ Fda,\ 1 & (13)\ \neg Fda,\ 1 \\
(14)\ d = a,\ 0 & (15)\ * \\
(16)\ * &
\end{array}
\end{array}
$$

6. Möjliga svar

Är det möjligt att undvika slutsatsen i detta argument? Så vitt jag kan se finns det i teorin sex möjliga svar på problemet i Avsnitt 5. Vi kan (i) Förkasta satslogiken (och/eller predikatlogiken); (ii) Överge (OD), $\neg(OA \wedge O\neg A)$; (iii) Ge upp antagandet att V_dFad; (iv) Förkasta uppfattningen att identitets-relationen är nödvändig; (v) Överge regeln (VE), att om B är logiskt ekvivalent med A och V_xA, så V_xB; eller (vi) Förkasta GR4. Låt oss kort undersöka dessa alternativ.

(i) Om man överger den klassiska satslogiken (och/eller den klassiska predikatlogiken) och t.ex. accepterar att det kan finnas sanna motsägelser, så kan man undvika slutsatsen att GR4 är falsk. Men det här tycks vara ett desperat förslag som man endast i yttersta nödfall bör ta till. Både den klassiska satslogiken och den klassiska predikatlogiken är extremt tilltalande och väletablerade.

(ii) Om vi förkastar OD kan vi undvika steg 9 i substitutionsargumentet ovan och kan då inte härleda en direkt motsägelse. Och det finns filosofer som har ifrågasatt denna princip. Personligen tycker jag dock att den är rimlig, och jag har försökt försvara den i andra arbeten (se Rönnedal (2012, ss. 73–96)). Även om man skulle överge (OD) så följer fortfarande steg 8 i argumentet. Och detta tycks vara en nästan lika orimlig konklusion som 10. Notera också att vi kan härleda ett liknande moraliskt dilemma för varje H, sådant att du vill att a utför H mot dig. Vidare gäller det att om du vill att a är ärlig mot dig, hjälper dig, håller sina löften till dig osv., så följer det att du *inte* bör vara ärlig mot a, *inte* bör hjälpa a, *inte* bör hålla dina löften till a osv. Och detta är orimligt, oavsett om resonemanget ger upphov till en motsägelse eller inte.

(iii) Om det inte är sant att du vill att a är ärlig mot dig, så följer inte våra problematiska slutsatser. Kanske vill inte alla personer att andra skall vara ärliga mot dem. Men det är nog ett rimligt antagande att det gäller för åtminstone nästan alla personer. Och för att argumentet skall gå igenom räcker det med att det finns en enda person som vill att någon annan är ärlig mot henne. Notera också att samma typ av argument kan användas även om F antas representera någon annan typ av handling. Så även om det skulle vara sant att det inte är fallet att det finns något x och något y sådana att x vill att y är ärlig mot x, så finns det säkert något x och något y sådana att x vill att y utför F mot x, för någon (problematisk typ av) handling F. Det här svaret tycks därför inte vara rimligt.

(iv) Om vi förkastar antagandet att identitetsrelationen är nödvändig, så kan vi inte längre visa steg 4 på samma sätt som ovan. Och det är inte uppenbart att all identitet är nödvändig. Om vi betraktar individkonstanterna d och a som rigida designatorer, om dessa refererar till samma personer i varje möjlig värld, och de refererar direkt till konkreta individer, följer det dock att deras icke-identitet är nödvändig. Och i vårt argument tycks det vara rimligt att anta att d och a är rigida designatorer och refererar direkt till dig och din partner. Om detta är riktigt, är (iv) inte ett svar som kan användas för att undvika substitutionsargumentet.

(v) Som vi redan har påpekat är regeln att V_xB följer ur V_xA om A och B är logiskt ekvivalenta knappast sann för alla personer i alla situationer. Men det är rimligt att anta att alla personer som är fullständigt rationella satisfierar denna princip. Och GR borde gälla för fullständigt rationella personer om några. Då har vi fortfarande ett problem.

(vi) Det enda rimliga alternativ som återstår tycks vara att överge GR4. Men om GR4 inte är en plausibel princip, betyder det också att den gyllene regeln inte är förnuftig? Innebär det att vi bör överge denna moraliska princip? Inte nödvändigtvis. Det finns andra tolkningar av GR. Vi nämnde några i Avsnitt 2. Vi skall nu emellertid undersöka ytterligare en interpretation. Betrakta följande läsning av GR.

> **(GR5).** Det gäller för alla individer x och y att: Om x vill att y utför handling H mot x, så bör x utföra H mot y, där ”H” står för en handling som kan uttryckas med vilket handlingspredikat som helst som inte innehåller några individkonstanter.

Låt $\underline{\sigma}$ vara en substitutionsfunktion som tar oss från predikat till formler som inte innehåller några individkonstanter. Då gäller det att den gyllene regeln tolkad som (GR5) medför (FGR) och alla satser som kan fås från (FGR) med hjälp av en substitutionsfunktion $\underline{\sigma}$, givet att (FGR) är fri för $\underline{\sigma}$. Om vi tolkar GR på detta sätt, så kan vi undvika problemet i Avsnitt 5. $\sigma H = (x_1 = a \wedge Fx_1x_2) \vee (\neg x_1 = a \wedge \neg Fx_1x_2)$. Men $(x_1 = a \wedge Fx_1x_2) \vee (\neg x_1 = a \wedge \neg Fx_1x_2)$ är en formel som innehåller individkonstanten a. σ kan därför inte tillämpas på (FGR) enligt GR5. Tolkar vi GR som GR5 är steg 3 i substitutionsargumentet inte tillåtet. Detta argument kan därför inte användas emot GR5.

GR5 är nästan lika stark som GR4. GR5 medför GR3, GR2 och GR1 och alla intuitivt riktiga konsekvenser av GR som vi nämnde i Avsnitt 2, dvs. (K1)–(K10) kan också härledas ur GR5 på samma sätt som vi gjorde i Avsnitt 4. Att GR5 är något svagare än GR4 tycks därför inte vara något allvarligt problem för en anhängare av den gyllene regeln.

Faktum är att det kan finnas andra skäl att föredra GR5 framför GR4. Begränsningen till substitutionsfunktioner vars värden inte innehåller några individkonstanter reflekterar väl den spridda uppfattningen att moralen *är* eller *bör vara* opartisk. Vad olika personer bör och inte bör göra och hur de bör och inte bör behandlas tycks inte vara beroende av vilka specifika individer de råkar vara, utan tycks endast bero på deras ”universella” egenskaper. Om GR tolkas på det här sättet är det möjligt att regeln kan härledas från en mer generell universaliserbarhets- eller opartiskhetsprincip. Att argumentera för detta ligger dock utanför den här uppsatsens ramar.

Konklusionen är att den gyllene regeln tolkad som GR5 inte kan vederläggas av substitutionsargumentet i Avsnitt 5 och att det förefaller vara helt rimligt att föredra GR5 framför GR4.

7. Slutsats

Enligt den gyllene regeln bör vi behandla andra så som vi själva vill bli behandlade. Jag har i den här uppsatsen undersökt några olika tolkningar av denna maxim och jag har framför allt intresserat mig för frågan om GR uttalar sig om *alla* handlingar eller inte och vad det innebär. Vi såg hur uttrycket ”alla handlingar” kan preciseras med hjälp av teorin för substitutionsfunktioner. Jag visade hur man med hjälp av tolkning 4 av GR kan härleda en mängd normer som alla tycks vara konsekvenser av den gyllene regeln. Enligt denna läsning är den gyllene regeln potentiellt mycket kraftfull och användbar. Vi såg emellertid att det s.k. substitutionsargumentet talar för att vår ursprungliga formulering, GR4, tycks vara alltför stark. Jag visade hur detta argument kan bemötas och hur man kan formulera en något svagare variant, GR5, som förefaller vara rimligare än GR4. Om resonemangen i den här uppsatsen är riktiga, så kan substitutionsargumentet inte användas för att vederlägga GR5. Och eftersom GR5 tycks vara en rimlig tolkning av den gyllene regeln, så innebär substitutionsargumentet inte något allvarligt problem för en anhängare av denna välkända princip. Det finns emellertid en mängd övriga uttryck i den gyllene regeln som också kan tolkas på många olika sätt. För att visa hur man kan bemöta olika potentiella invändningar, bör man nog säga något mer också om dessa uttryck. Jag hoppas kunna återvända till ämnet vid något annat tillfälle.[6]

Referenser

Blackstone, W. T. (1965). The Golden Rule: A Defense. *Southern Journal of Philosophy*, ss. 172–177.

Bruton, S. V. (2004). Teaching the Golden Rule. *Journal of Business Ethics*, Vol. 49, Nr. 2, ss. 179–187.

Cadoux, A. T. (1912). The Implications of the Golden Rule. *International Journal of Ethics*, Vol. 22, Nr. 3, ss. 272–287.

Carson, T. L. (2010). *Lying and Deception: Theory and Practice*. Oxford: Oxford University Press.

Carson, T. L. (2013). Golden Rule. I Hugh LaFollette (red.) *The International Encyclopedia of Ethics*, ss. 2186–2192.

Duxbury, N. (2009). Golden Rule Reasoning, Moral Judgement and Law. *Notre Dame Law Review* 84, ss. 1529–1605.

[6] I Rönnedal (2016) diskuterar jag några potentiella problem med vissa formuleringar av den gyllene regeln. Jag visar också hur dessa problem kan bemötas.

Gensler, H. J. (1986). Ethics is Based on Rationality. *The Journal of Value Inquiry* 20, ss. 251–264.

Gensler, H. J. (1996). *Formal Ethics*. London and New York: Routledge.

Gensler, H. J. (2013). *Ethics and the Golden Rule*. New York and London: Routledge.

Gewirth, A. (1978). The golden rule rationalized. *Midwest Studies in Philosophy*, I11, ss. 133–147.

Gould, J. A. (1980). Blackstone's Meta-Not-So-Golden-Rule. *The Southern Journal of Philosophy*, Vol. 18, Issue 4, ss. 509–513.

Hare, R. M. (1963). *Freedom and Reason*. Oxford: Oxford University Press.

Hertzler, J. O. (1934). On Golden Rules. *International Journal of Ethics*, Vol. 44, Nr. 4, ss. 418–436.

Hirst, E. W. (1934). The Categorical Imperative and the Golden Rule. *Philosophy*, Vol. 9, Nr. 35, ss. 328–335.

Hobbes, T. (1985). *Leviathan*. Penguin Books. (red. C. B. Macpherson). (Ursprungligen publicerad 1651.)

Hoche, H.-U. (1978). Die Goldene Regel. Neue Aspekte eines alten Moralprinzips. *Zeitschrift für philosophische Forschung*, Bd. 32, H. 3, ss. 355–375.

Huang, Y. (2005). A Copper Rule versus the Golden Rule: A Daoist-Confucian Proposal for Global Ethics. *Philosophy East and West*, Vol. 55, Nr. 3, ss. 394–425.

Kleene, S. C. (1971). *Introduction to Metamathematics*. Groningen: Wolters-Noordhoff Publishing.

Mill, J. S. (1987). *Utilitarianism*. Buffalo, New York: Prometheus Books. (Ursprungligen publicerad 1863).

Neusner, J. och Chilton, B. (red.) (2008). *The Golden Rule: The Ethics of Reciprocity in the World Religions*. Continuum.

Pufendorf, S. (1964). *On the Law of Nature and Nations*. New York: Wildy and Sons. (Ursprungligen publicerad 1672).

Reinikainen, J. (2005). The Golden Rule and the Requirement of Universalizability. *The Journal of Value Inquiry* 39, ss. 155–168.

Rönnedal, D. (2012). *Extensions of Deontic Logic: An Investigation into some Multi-Modal Systems*. Department of Philosophy, Stockholm University.

Rönnedal, D. (2015). The Golden Rule and The Platinum Rule. *The Journal of Value Inquiry*, Volume 49, Issue 1, ss. 221–236.

Rönnedal, D. (2016). Den Gyllene Regeln och Intra- och Interpersonella Viljekonflikter. *Filosofiska Notiser*, Årgång 3, Nr 2, Augusti, ss. 81–106.

Schurz, G. (1997). *The Is-Ought Problem: An Investigation in Philosophical Logic*. Springer.

Singer, M. G. (1963). The Golden Rule. *Philosophy*, Vol. 38, Nr. 146, ss. 293–314.

Wattles, J. (1996). *The Golden Rule*. New York, Oxford: Oxford University Press.

Weiss, P. (1941). The Golden Rule. *The Journal of Philosophy*, Vol. 38, Nr. 16, ss. 421–430.

Daniel Rönnedal
Filosofiska institutionen
Stockholms universitet
daniel.ronnedal@philosophy.su.se