Filosofiska Notiser

Årgång 2, Nr 1, Februari 2015

ISSN: 2002-0198

Hemsida: www.filosofiskanotiser.com

Fritt Val Tillåtelser

Daniel Rönnedal

Abstrakt
Den här uppsatsen handlar om fritt val (FV) tillåtelser (FVT). Jag går igenom den s.k. fritt val tillåtelser paradoxen och nämner några möjliga lösningar på denna. Därefter presenterar jag mitt eget förslag på hur man bör förstå tillåtelser av detta slag och hur man kan lösa (FVT) paradoxen. Jag tar upp några potentiella invändningar mot denna analys och visar hur dessa kan bemötas. Ibland har (FVT) paradoxen använts som ett argument emot s.k. standard deontisk logik (SDL). Jag argumenterar för att man kan acceptera förekomsten av (FV) tillåtelser utan att behöva förkasta (SDL). Däremot pekar diskussionen på behovet av en kvantifierad deontisk logik.

1. Introduktion

Betrakta följande resonemang.

> Argument 1
> Du får ta en bulle eller en syltmunk.
> Alltså får du ta en bulle och du får ta en syltmunk.

Detta argument tycks vara giltigt. Slutsatsen tycks följa ur premissen. Antag att du är på besök hos din farmor och farfar. Farmor sträcker fram ett fat med flera bullar och munkar på, och säger: ”Du får ta en bulle eller en syltmunk”. I denna situation verkar det vara rimligt för dig att dra slutsatsen att du får ta en bulle och att du får ta en syltmunk (men inte nödvändigtvis att du får ta både en bulle och en syltmunk).

Följande symbolisering av detta resonemang i s.k. standard deontisk logik (SDL) är emellertid inte giltigt.[1]

$$P(B \vee S)$$
$$PB \wedge PS$$

[1] För mer information om SDL och deontisk logik, se t.ex. Gabbay, Horty, Parent, van der Meyden, & van der Torre (red.) (2013), Hilpinen (red.) (1971), (1981), Rönnedal (2010), eller Åqvist (1987), (2002).

Där B står för ”Du tar en bulle”, S för ”Du tar en syltmunk” och P är en satsoperator som utläses ”Det är tillåtet att”, eller ”Det får vara fallet att”. Följande modell bevisar detta. Mängden av alla möjliga världar består av w_0 och w_1; w_1 är deontiskt tillgänglig från w_0; B är falsk i w_1 och S är sann i w_1. För att förstärka denna modell till en OS5+ modell kan vi anta att w_1 också är tillgänglig från sig själv.

Men hur skall vi då förklara att argument 1 tycks vara giltigt, att det tycks vara rimligt att härleda slutsatsen ur premissen?

Ett möjligt svar på denna fråga är att hävda att SDL är ett för svagt logiskt system, att vi bör lägga till någon premiss till SDL som gör argument 1 giltigt.

Betrakta följande princip, fritt val principen (FVP).

(FVP) $P(A \lor B) \to (PA \land PB)$

Om vi lägger till FVP till SDL, kan vi enkelt visa att slutsatsen i argument 1 kan härledas från premissen i detta argument. Gör vi det, följer emellertid ett antal kontraintuitiva slutsatser (Hansson (2013)). Låt oss titta närmare på några av dessa. (Omedelbart efter den problematiska satsen i varje exempel har jag angivit vilka teorem och regler som används i härledningarna.)

Problem 1 (Kamp 1973). $Op \to O(p \land q)$
P-Extensionalitet (P-Ext): Om $A \leftrightarrow B$ är ett teorem, så är också $PA \leftrightarrow PB$ ett teorem. Definierbarhet (Def-O): $OA \leftrightarrow \neg P \neg A$. SL innebär att steget följer med vanlig satslogik.

1. $P(\neg p \lor \neg q) \to P\neg p$ [FVP, SL]
2. $P\neg(p \land q) \to P\neg p$ [1, P-Ext, SL]
3. $\neg P\neg p \to \neg P\neg(p \land q)$ [2, SL]
4. $Op \to O(p \land q)$ [3, Def-O]

Men $Op \to O(p \land q)$ är en orimlig princip. Här är en kontraintuitiv instans: Om det är obligatoriskt att du betalar skatt, så är det obligatoriskt att du betalar skatt och bränner ner skatteverkets kontor.

Problem 2 (von Wright 1968). $Op \to Pq$
Axiom (D): $OA \to PA$. O-försvagning (OF): $OA \to O(A \lor B)$

1. $P(p \lor q) \to Pq$ [FVP, SL]
2. $O(p \lor q) \to P(p \lor q)$ [D]
3. $Op \to O(p \lor q)$ [OF]
4. $Op \to Pq$ [1-3, SL]

Men $Op \to Pq$ är en orimlig princip, som hävdar att om någonting bör vara fallet, så är allt tillåtet. Enligt denna sats gäller det t.ex. att om du bör skänka pengar till något välgörande ändamål, så är det tillåtet att du misshandlar din partner.

Problem 3 (Makinson 1984): $Pp \rightarrow Pq$.
O-Extensionalitet (O-Ext): Om $A \leftrightarrow B$ är ett teorem, så är också $OA \leftrightarrow OB$ ett teorem. O-distribution (OD): $O(A \wedge B) \rightarrow (OA \wedge OB)$. (Def-P): $PA \leftrightarrow \neg O\neg A$.

1. $P(p \vee q) \rightarrow Pq$ [FVP, SL]
2. $O(\neg p \wedge \neg q) \rightarrow O\neg p$ [OD, SL]
3. $O\neg(p \vee q) \rightarrow O\neg p$ [2, O-Ext, SL]
4. $\neg O\neg p \rightarrow \neg O\neg(p \vee q)$ [3, SL]
5. $Pp \rightarrow P(p \vee q)$ [4, Def-P]
6. $Pp \rightarrow Pq$ [1, 5, SL]

Men $Pp \rightarrow Pq$ är en orimlig princip. Den säger att om något är tillåtet, så är allt tillåtet. T.ex.: Om det är tillåtet att du lämnar landet, så är det tillåtet att du stjäl din grannes bil.

Problem 4 (Hilpinen 1982): $Pp \rightarrow P(p \wedge q)$.
(P-Ext).

1. $P((p \wedge q) \vee (p \wedge \neg q))$
   $\rightarrow (P(p \wedge q) \wedge P(p \wedge \neg q))$ [FVP]
2. $Pp \rightarrow (P(p \wedge q) \wedge P(p \wedge \neg q))$ [1, P-Ext, SL]
3. $Pp \rightarrow P(p \wedge q)$ [2, SL]

Notera att detta är ett problem inte bara för SDL och för alla s.k. normala deontiska system, utan för alla deontiska system som innehåller (Ext), t.ex. alla s.k. klassiska system. $Pp \rightarrow P(p \wedge q)$ är en kontraintuitiv princip. Här är en problematisk instans: Om det är tillåtet att du gifter dig, så är det tillåtet att du gifter dig och spränger slottet i luften.

Problem 5. $Op \rightarrow Oq$.
(P-Ext). (Def-O). (OD).

1. $P(\neg p \vee \neg q) \rightarrow P\neg p$ [FVP, SL]
2. $P\neg(p \wedge q) \rightarrow P\neg p$ [1, P-Ext, SL]
3. $\neg P\neg p \rightarrow \neg P\neg(p \wedge q)$ [2, SL]
4. $Op \rightarrow O(p \wedge q)$ [3, Def-O]
5. $O(p \wedge q) \rightarrow Oq$ [OD, SL]
6. $Op \rightarrow Oq$ [4, 5, SL]

$Op \rightarrow Oq$ säger att om något är obligatoriskt, så är allting obligatoriskt. Detta är orimligt. Här är ett exempel på en kontraintuitiv instans. Om det är obligatoriskt att du går ut grundskolan, så är det obligatoriskt att du startar tredje världskriget.

Faktum är att vi kan härleda en motsägelse i SDL om vi lägger till FVP. Antag att vi utvidgar SDL med FVP. Vi har sett att vi då kan härleda $Pp \rightarrow Pq$. Vi kan byta ut p och q mot vilka satser som helst. Alltså $P\top \rightarrow P\bot$. Men i SDL har vi $P\top$. Alltså $P\bot$. Alla normala deontiska system innehåller emellertid $\neg P\bot$. Detta är absurt.

Dessa problem visar på ett tydligt sätt att det är orimligt att acceptera alla satser och regler i SDL om det är rimligt att acceptera FVP. Omvänt gäller det att om vi har goda skäl att hålla fast vid SDL, så har vi goda skäl att förkasta FVP. Och vi tycks ha goda skäl att acceptera SDL. Samtidigt är FVP en intuitivt rimligt princip. Satsen ”Du får ta en bulle och du får ta en syltmunk” tycks följa ur satsen ”Du får ta en bulle eller en syltmunk”.

Hur skall man ställa sig till detta problem, som vi skall kalla ”fritt val tillåtelser (FVT) paradoxen”?

2. Möjliga lösningar

Många olika möjliga lösningar på fritt val tillåtelser paradoxen har presenterats. Ett förslag går ut på att införa en ny ”fritt val” operator som satisfierar FVP eller att helt enkelt symbolisera ”Du får A eller B” som $PA \wedge PB$. Se t.ex. Föllesdal & Hilpinen (1971), von Wright (1971), och Woleński (1980).

Ett annat förslag går ut på att förkasta SDL och utveckla någon alternativ logik. Se t.ex. Dignum, Meyer & Wieringa (1996), som använder en dynamisk deontisk logik (se också Gabbay, Gammaitoni & Sun (2014)), Asher & Bonevac (2005), som utvecklar en deontisk logik baserad på en icke-monotonisk logik, Barker (2010), som utgår ifrån en deontisk logik baserad på ”linjär logik”, och Aher (2012), som introducerar en deontisk logik baserad på en s.k. ”inkvisitiv” (eng: inquisitive) semantik.

En alternativ lösning hävdar att slutsatsen i argument 1 (och liknande argument) inte följer (semantiskt) från premissen. FVP är inte en giltig princip; det handlar om en *pragmatisk* implikation. Premissen i argument 1 (och liknande argument) implicerar *pragmatiskt* slutsatsen, men premissen kan vara sann samtidigt som slutsatsen är falsk. Se t.ex. Zimmerman (2000), Schulz (2005b), Fox (2007), och Franke (2010).

Ytterligare ett alternativ går ut på att införa en dold Endast-operator som inte explicit uttrycks men förstås av språkanvändarna. Kamp (1973) tycks vara den första som föreslår en lösning av detta slag. Flera andra har försökt utveckla denna idé. Se t.ex. Dignum, Meyer & Wieringa (1996).

Några filosofer har menat att ”fri tillåtelse” inte är en egenskap hos en disjunktion utan en egenskap hos mängden av disjunkter. Se t.ex. Hansson (2001), särskilt ss. 130-131, och Hansson (2013).

Slutligen kan man hävda att problemen med fritt val tillåtelser och fritt val tillåtelser paradoxen uppstår som en följd av en ”felaktig” eller ”dålig” översättning av de naturliga satserna i argument 1 (och liknande argument) till formella satser. Slutsatsen följer, om vi symboliserar dessa satser ”korrekt”. En lösning av den här typen finns antydd i Parks (1973). Både Stenius (1982) och Makinson (1984) argumenterar explicit på detta sätt (se också Åqvist (1965)). Den ”lösning” jag argumenterar för i den här uppsatsen är en utveckling av idéer som presenteras i Stenius (1982) och i Makinson (1984).[2]

3. Lösningsförslag 1

Enligt den lösning av fritt val tillåtelser paradoxen som jag föreslår i den här uppsatsen behöver vi inte överge SDL eller några teorem eller regler i detta system. Det går att tolka argument 1 på ett sådant sätt att det blir semantiskt giltigt. Det är alltså inte nödvändigtvis så att premissen endast *pragmatiskt* implicerar slutsatsen. Lösningen förutsätter dock att vi utvidgar SDL med predikatlogik. Jag kommer att använda mig av en kvantifierad temporal aletisk-deontisk logik för att symbolisera flera olika fritt val tillåtelser. De tekniska detaljerna i dessa system utvecklas i Rönnedal (2014).[3] Jag kommer i det här avsnittet att presentera ett lösningsförslag. På grund av vissa problem med detta (se avsnitt 5) kommer jag att undersöka ett nytt, modifierat förslag i avsnitt 6. Först skall vi emellertid analysera ett antal enklare former av fritt val tillåtelser.

Betrakta följande scenario. Du är på besök hos din farmor och farfar. Farmor håller fram ett fat med kakor och säger: ”Du får ta en kaka”. Vad betyder satsen ”Du får ta en kaka”? Vi bör skilja mellan åtminstone fyra olika möjliga tolkningar.

1. Det finns en kaka som det är tillåtet att du tar.
2. Det är tillåtet att det finns en kaka som du tar.
3. Du får ta vilken kaka som helst.
4. Du får ta alla kakor.

Sats 3 tycks även kunna uttryckas på följande sätt: ”För varje kaka gäller det att du får tar den”. Låt Kx vara ett predikat som läses ”x är en kaka [på fatet]”, Txy ett predikat som läses ”x tar y” och d en singulär term som refererar till dig. ∃x är en partikulär kvantifikator som läses ”Det finns ett x sådant att” och ∀x en universell kvantifikator som läses ”Det gäller för alla x

[2] För mer information om fritt val tillåtelser, se Aloni (2007), Anglberger, Dong & Roy (2014), Chemla (2009), Geurts (2005), Geurts & Pouscoulous (2009), Giannakidou (2001), Jennings (1985), Merin (1992), Nute (1985), Saeboe (2001), Schulz (2005), Simons (2005), van Rooij (2006), och Åqvist (1965), (1987), särskilt ss. 47-55.

[3] Se också Rönnedal (2012) och (2012b). Notera att jag använder något andra symboler i den här uppsatsen.

att”. $\exists$ och $\forall$ kan antingen tolkas possibilistiskt eller aktualististiskt beroende på kontexten. Då kan satserna 1-4 symboliseras på följande sätt i en kvantifierad deontisk logik.

1′. $\exists x(Kx \wedge PTdx)$. Det finns ett x sådant att x är en kaka och det är tillåtet att du tar x.
2′. $P\exists x(Kx \wedge Tdx)$. Det är tillåtet att det finns ett x sådant att x är en kaka och du tar x.
3′. $\forall x(Kx \rightarrow PTdx)$. Det gäller för alla x att om x är en kaka, så är det tillåtet att du tar x.
4′. $P\forall x(Kx \rightarrow Tdx)$. Det är tillåtet att det gäller för alla x att om x är en kaka, så tar du x.

Hur förhåller sig dessa formler till varandra? I vanlig kvantifierad deontisk logik följer $\exists x(Kx \wedge PTdx)$ ur $\forall x(Kx \rightarrow PTdx)$ om vi antar att det finns åtminstone en kaka ($\exists xKx$). I övrigt är de olika satserna logiskt oberoende av varandra, förutsatt att vi inte gör några särskilda antaganden. 1′ medför inte 2′; 2′ medför inte 3′ ; 3′ medför inte 4′; 4′ medför inte 3′; 3′ medför inte 2′; 2′ medför inte 1′ osv. Låt oss sammanfatta detta i ett teorem.

Teorem 1. (i) (3′) $\forall x(Kx \rightarrow PTdx)$ och $\exists xKx$ medför (1′) $\exists x(Kx \wedge PTdx)$. **(ii)** (1′) $\exists x(Kx \wedge PTdx)$, (2′) $P\exists x(Kx \wedge Tdx)$, (3′) $\forall x(Kx \rightarrow PTdx)$, och (4′) $P\forall x(Kx \rightarrow Tdx)$ är alla logiskt oberoende av varandra. Dvs. (1′) medför inte (2′) och (2′) medför inte (1′), (1′) medför inte (3′) och (3′) medför inte (1′), osv. Detta gäller i alla (”normala”) kvantifierade deontiska system om vi inte gör några särskilda antaganden.

Bevis. Lämnas till läsaren. ■

Vilken av dessa tolkningar är rimligast? Det tycks vara ganska uppenbart att ”Du får ta en kaka” inte betyder samma sak som ”Du får ta alla kakor”, det är möjligt att du får ta en kaka även om du inte får ta alla kakor. Däremot är det inte uppenbart huruvida ”Du får ta en kaka” betyder samma sak som 1′, 2′, eller 3′ eller om det finns någon semantisk skillnad mellan 1′ och 2′. Enligt mina intuitioner används ”Du får ta en kaka” ofta för att säga samma sak som ”Du får ta en kaka, vilken som helst”, en sats som i sin tur säger samma sak som ”Du får ta vilken kaka som helst”. Tolkning 1′ och 2′ är också möjliga, men ofta mindre intuitiva. Enligt 1′ finns det t.ex. en specifik kaka som du får ta. Detta medför inte att du får ta vilken kaka som helst.

Betrakta följande scenario. Antag att det finns 3 kakor på fatet: en syltkaka, en pepparkaka och en fruktkaka. Antag också att farmor säger: ”Du får ta en kaka, med du får inte ta syltkakan och du får inte ta fruktkakan”. Detta tycks inte vara inkonsistent, men är samtidigt ett underligt yttrande. Om farmor vill säga att du får ta pepparkakan, men ingen annan kaka, varför säger hon då inte bara ”Du får ta pepparkakan”? Mer generellt: antag att du får ta en kaka och att det innebär att det finns en specifik kaka som du får ta, samtidigt som det gäller att du inte får ta någon annan kaka. Då förefaller det vara mer naturligt att säga: ”Du får ta den där kakan”, och samtidigt peka på den kaka man menar, istället för ”Du får ta en kaka”.

Enligt denna tolkning säger alltså ”Du får ta en kaka” samma sak som ”Du får ta en kaka, vilken som helst” eller ”Du får ta vilken kaka som helst”. Detta kan också, något mer krystat, uttryckas på följande sätt: ”En kaka, vilken som helst, får du ta”, ”Vilken kaka som helst får du ta” eller ”En kaka får du ta”. Dessa satser kan symboliseras på följande sätt: $\forall x(Kx \rightarrow PTdx)$.

Mitt förslag går ut på att argument 1 (och analoga argument) kan symboliseras på liknande sätt. Argument 1 kan t.ex. ges följande läsning:

> Argument 1′
> Du får ta en bulle, vilken som helst, eller en syltmunk,
> vilken som helst.
> Alltså får du ta en bulle, vilken som helst, och du får ta en syltmunk,
> vilken som helst.

Eller ekvivalent

> Argument 1″
> Du får ta vilken bulle eller (vilken) syltmunk som helst.
> Alltså får du ta vilken bulle som helst och du får ta vilken syltmunk
> som helst.[4]

Följande symbolisering av argument 1 är naturlig givet denna tolkning.

$$\forall x((Bx \vee Sx) \rightarrow PTdx)$$
$$\forall x(Bx \rightarrow PTdx) \wedge \forall x(Sx \rightarrow PTdx)$$

Premissen kan läsas på följande vis: ”Det gäller för alla x att om x är en bulle eller (x är) en syltmunk, så är det tillåtet att du tar x”, och slutsatsen på följande vis: ”Det gäller för alla x att om x är en bulle så är det tillåtet att du tar x och det gäller för alla x att om x är en syltmunk så är det tillåtet att du

[4] Här följer ett par ”osmidiga” varianter, som emellertid blottlägger argumentets struktur bättre. Premiss: En bulle (vilken som helst) eller en syltmunk (vilken som helst) får du ta. Slutsats: Alltså, en bulle (vilken som helst) får du ta och en syltmunk (vilken som helst) får du ta. Eller. Premiss: Vilken bulle eller (vilken) syltmunk som helst får du ta. Slutsats: Alltså, vilken bulle som helst får du ta och vilken syltmunk som helst får du ta.

tar x”. Detta argument är giltigt av rent predikatlogiska skäl, vilket följande semantiska tablå bevisar.

$$
\begin{array}{c}
\forall x((Bx \lor Sx) \to PTdx), w_0t_0 \\
\neg(\forall x(Bx \to PTdx) \land \forall x(Sx \to PTdx)), w_0t_0 \\
\swarrow \qquad \searrow \\
\begin{array}{cc}
\begin{array}{c}
\neg\forall x(Bx \to PTdx), w_0t_0 \\
\exists x\neg(Bx \to PTdx), w_0t_0 \\
\neg(Bc \to PTdc), w_0t_0 \\
Bc, w_0t_0 \\
\neg PTdc, w_0t_0 \\
(Bc \lor Sc) \to PTdc, w_0t_0 \\
\swarrow \quad \searrow \\
\begin{array}{ll}
\neg(Bc \lor Sc), w_0t_0 & PTdc, w_0t_0 \\
\neg Bc, w_0t_0 & * \\
\neg Sc, w_0t_0 & \\
* &
\end{array}
\end{array}
&
\begin{array}{c}
\neg\forall x(Sx \to PTdx), w_0t_0 \\
\exists x\neg(Sx \to PTdx), w_0t_0 \\
\neg(Sc \to PTdc), w_0t_0 \\
Sc, w_0t_0 \\
\neg PTdc, w_0t_0 \\
(Bc \lor Sc) \to PTdc, w_0t_0 \\
\swarrow \quad \searrow \\
\begin{array}{ll}
\neg(Bc \lor Sc), w_0t_0 & PTdc, w_0t_0 \\
\neg Bc, w_0t_0 & * \\
\neg Sc, w_0t_0 & \\
* &
\end{array}
\end{array}
\end{array}
\end{array}
$$

Låt oss undersöka om den här typen av lösning kan generaliseras till ett antal andra varianter av argument 1. Antag att det bara finns *en* bulle och *en* syltmunk på fatet. Då tycks följande argument vara giltigt.

Argument 2
Du får ta bullen eller syltmunken.
Det följer att du får ta bullen och att du får ta syltmunken.

En naturlig symbolisering av detta argument i kvantifierad deontisk logik ser ut på följande sätt:

$P(Tdb \lor Tds)$
$PTdb \land PTds$

Där b refererar till bullen och s till syltmunken. Men detta argument är inte giltigt. Om vi däremot använder följande symbolisering, förhåller det sig annorlunda.

$\forall x((x = b \lor x = s) \to PTdx)$
$PTdb \land PTds$

Enligt denna tolkning säger premissen: ”Det gäller för alla x att om x är identisk med bullen eller (x är identisk) med syltmunken, så är det tillåtet att du tar x”. Detta argument är giltigt i SDL med predikatlogik, vilket bevisas av följande semantiska tablå.

$$\begin{array}{c}
\forall x((x = b \vee x = s) \rightarrow PTdx),\ w_0t_0 \\
\neg(PTdb \wedge PTds),\ w_0t_0 \\
(b = b \vee b = s) \rightarrow PTdb,\ w_0t_0 \\
(s = b \vee s = s) \rightarrow PTds,\ w_0t_0 \\
\swarrow \qquad\qquad\qquad \searrow
\end{array}$$

$$\begin{array}{l}
\neg(b = b \vee b = s),\ w_0t_0 \\
\neg b = b,\ w_0t_0 \\
\neg b = s,\ w_0t_0 \\
*
\end{array}
\qquad
\begin{array}{c}
PTdb,\ w_0t_0 \\
\swarrow \qquad \searrow
\end{array}$$

$$\begin{array}{l}
\neg(s = b \vee s = s),\ w_0t_0 \\
\neg s = b,\ w_0t_0 \\
\neg s = s,\ w_0t_0 \\
*
\end{array}
\qquad
\begin{array}{l}
PTds,\ w_0t_0 \\
sw_0w_1t_0 \\
Tdb,\ w_1t_0 \\
sw_0w_2t_0 \\
Tds,\ w_2t_0 \\
\swarrow \qquad \searrow
\end{array}$$

$$\begin{array}{l}
\neg PTdb,\ w_0t_0 \\
O\neg Tdb,\ w_0t_0 \\
\neg Tdb,\ w_1t_0 \\
*
\end{array}
\qquad
\begin{array}{l}
\neg PTds,\ w_0t_0 \\
O\neg Tds,\ w_0t_0 \\
\neg Tds,\ w_2t_0 \\
*
\end{array}$$

Rent spontant kan symboliseringen av premissen i detta argument förefalla något långsökt. Men i ljuset av vår formalisering av argument 1, tycks den ändå vara ganska naturlig och ger ett intuitivt rimligt resultat.

Följande utvikning kanske också kan stödja denna uppfattning. Antag att det finns exakt en bulle på fatet och exakt en munk på fatet och ingenting annat. Hur uttrycker man att du får ta ett av dessa bakverk, men inte nödvändigtvis båda (om man vill undvika den otympliga satsen ”Det är tillåtet att du tar bullen och det är tillåtet att du tar munken”)? Det tycks finnas tre möjligheter.

1. Du får ta bullen och munken.
2. Du får ta bullen eller du får ta munken.
3. Du får ta bullen eller munken.

1 är för stark eftersom den innebär att du får ta både bullen och munken, och det är inte det vi menar. 2 är för svag eftersom den är förenlig med att du inte får ta bullen (och också med att det är förbjudet att du tar munken), även om den inte är förenlig med att det är förbjudet att du tar bullen *och* förbjudet att du tar munken. 3 tycks vara det enda alternativ som återstår. Men 3 antyder att fritt val tillåtelser har den logiska formen $P(A \vee B)$, vilket är problematiskt eftersom detta är ekvivalent med $PA \vee PB$. Men om resonemangen hittills är

korrekta, så kan innebörden av 3 tolkas på följande sätt: ”För vadhelst (som finns på fatet) som är identiskt antingen med bullen eller med munken gäller det att du får ta det”, eller: ”För varje x (som finns på fatet) sådant att x är identiskt med bullen eller med munken gäller det att det är tillåtet att du tar x”, vilket var precis vad vi ville uttrycka. ”Du får ta bullen eller munken, men inte båda” kan symboliseras på följande sätt: $\forall x((x = b \vee x = m) \rightarrow PTdx) \wedge \neg P(Tdb \wedge Tdm)$. Från detta följer $PTdm \wedge PTdb$. Men $PTdm \wedge PTdb$ medför inte $P(Tdb \wedge Tdm)$; $\{\forall x((x = b \vee x = m) \rightarrow PTdx) \wedge \neg P(Tdb \wedge Tdm)\}$ är konsistent.

Låt oss betrakta ytterligare ett exempel.

Argument 3 (Kamp (1973))
Du får gå till stranden eller på bio.
Alltså får du gå till stranden och du får gå på bio.

Jag föreslår att detta argument kan tolkas på två sätt beroende på om vi antar att ”stranden” och ”bio” refererar till en specifik strand och en specifik biograf eller ”generiskt” till någon strand och någon biograf. (I avsnitt 9 nämner jag ytterligare en tolkning.)

Argument 3′
Du får gå till stranden eller till biografen.
Alltså får du gå till stranden och du får gå till biografen.

Argument 3″
Du får gå till en strand eller till en biograf.
Alltså får du gå till en strand och du får gå till en biograf.

Låt Gxy stå för ”x går till y”, s referera till stranden, och b till biografen. Då kan argument 3′ symboliseras på följande sätt: Premiss: $\forall x((x = s \vee x = b) \rightarrow PGdx)$. Slutsats: $\forall x(x = s \rightarrow PGdx) \wedge \forall x(x = b \rightarrow PGdx)$. Premissen läses ”Det gäller för alla x att om x är lika med stranden eller (x är lika med) biografen, så är det tillåtet att du går till x”. Slutsatsen läses ”Det gäller för alla x att om x är identisk med stranden så är det tillåtet att du går till x och det gäller för alla x att om x är identisk med biografen så är det tillåtet att du går till x”.

Argument 3″ kan i sin tur tolkas på följande sätt:

Argument 3‴
Du får gå till en strand (vilken som helst) eller till en biograf (vilken som helst). Dvs. Du får gå till vilken strand som helst eller till vilken biograf som helst.

> Alltså får du gå till en strand, vilken som helst, och du får gå till en biograf, vilken som helst. Dvs. Du får gå till vilken strand som helst och du får gå till vilken biograf som helst.

Låt Sx stå för x är en strand, och Bx för x är en biograf. Då kan argument 3‴ symboliseras på följande sätt. Premiss: $\forall x((Sx \vee Bx) \rightarrow PGdx)$. Slutsats: $\forall x(Sx \rightarrow PGdx) \wedge \forall x(Bx \rightarrow PGdx)$. Premissen läses ”Det gäller för alla x att om x är en strand eller (x är) en biograf, så är det tillåtet att du går till x”. Slutsatsen läses ”Det gäller för alla x att om x är en strand så är det tillåtet att du går till x och det gäller för alla x att om x är en biograf så är det tillåtet att du går till x.

Det bör vid det här laget vara tämligen uppenbart hur denna typ av analys kan generaliseras.[5]

4. Problem 1: Tillbakadragna slutsatser

Låt oss nu undersöka några möjliga problem med denna lösning.

Betrakta följande scenario. Du befinner dig på en restaurang och har ätit färdigt huvudrätten. Servitrisen säger då: ”Du får ta en tårtbit eller en bakelse (till efterrätt). Men jag vet inte vilket, jag är ny här. Låt mig kolla med köket”. I den här situationen förefaller vi inte kunna dra slutsatsen att du får ta en tårtbit och att du får ta en bakelse från påståendet att du får ta en tårtbit eller en bakelse. Följande argument tycks alltså inte vara giltigt.

> Argument 4
> Du får ta en tårtbit eller en bakelse.
> Alltså får du ta en tårtbit och du får ta en bakelse.

Men om vi symboliserar detta argument på samma sätt som vi symboliserade argument 1, så blir resultatet giltigt. Följande formalisering av argument 4 är giltig. Premiss: $\forall x((Tx \vee Bx) \rightarrow PTdx)$. Slutsats: $\forall x(Tx \rightarrow PTdx) \wedge \forall x(Bx \rightarrow PTdx)$. Vad bör vi säga om den här invändningen mot vår analys?

Ett enkelt svar går ut på att vi bör symbolisera detta argument på ett sätt som tar hänsyn till den kontext i vilken premissen yttras, t.ex. på följande sätt. Premiss: $P(T \vee B)$ $[P(\exists x(Tx \wedge Tdx) \vee \exists x(Bx \wedge Tdx))]$, slutsats: $PT \wedge PB$

---

[5] Här följer emellertid en lista på ytterligare argument som tycks kunna symboliseras på liknande sätt. Du får skriva en uppsats om Hume eller (om) Kant. Alltså får du skriva en uppsats om Hume och du får skriva en uppsats om Kant (Hansson 2013). Du får skriva en uppsats om en tysk (filosof) eller en engelsk filosof. Alltså får du skriva en uppsats om en tysk filosof och du får skriva en uppsats om en engelsk filosof. Du får framföra en sång eller en dans. Alltså får du framföra en sång och du får framföra en dans. Du får ta ett äpple eller ett päron. Alltså får du ta ett äpple och du får ta ett päron (Merin 1992, van Rooy 2000, Schulz 2005b). Jane får besöka Henry eller Matilda. Alltså får Jane besöka Henry och Jane får besöka Matilda (Simons 2005).

$[P\exists x(Tx \land Tdx) \land P\exists x(Bx \land Tdx)]$, där T står för ”Du tar en tårtbit” och B står för ”Du tar en bakelse” [Tx: x är en tårtbit, Bx: x är en bakelse, Txy: x tar y]. Från $P(T \lor B)$ $[P(\exists x(Tx \land Tdx) \lor \exists x(Bx \land Tdx))]$ följer $PT \lor PB$ $[P\exists x(Tx \land Tdx) \lor P\exists x(Bx \land Tdx)]$ i SDL. Men $PT \land PB$ $[P\exists x(Tx \land Tdx) \land P\exists x(Bx \land Tdx)]$ är, som vi redan har påpekat, inte härledbar från $P(T \lor B)$ $[P(\exists x(Tx \land Tdx) \lor \exists x(Bx \land Tdx))]$. Enligt denna tolkning är argument 4 inte giltigt. Däremot kan vi sluta oss till att det är tillåtet att du tar en tårtbit eller tillåtet att du tar en bakelse. Och detta tycks stämma väl överens med våra intuitioner. Dvs. vi tycks kunna sluta oss till att det är tillåtet att du tar en tårtbit eller tillåtet att du tar en bakelse i vårt scenario, men inte att det är tillåtet att du tar en tårtbit och att det är tillåtet att du tar en bakelse.

Vissa problem kvarstår dock. Hur vet vi t.ex. när vi skall symbolisera ett argument på det ena eller det andra sättet? Kan vi säga något om detta?

Låt oss gå tillbaka till satserna 1-4 och se hur dessa skulle kunna symboliseras i ett kvantifierat temporalt aletiskt-deontiskt system där vi antar att det förflutna och nuet är determinerat, men där framtiden kan vara öppen. I ett sådant system rör alla praktiska normer framtiden (se Rönnedal (2012)). $\underline{F}$ är en satsoperator som läses ”Det kommer någon gång i framtiden vara fallet att”. Då formaliserar vi 1-4 på följande sätt.

> $1''$. $\exists x(Kx \land P\underline{F}Tdx)$. Det finns ett x sådant att x är en kaka och det är tillåtet att du tar x (någon gång i framtiden).
> $2''$. $P\exists x(Kx \land \underline{F}Tdx)$. Det är tillåtet att det finns ett x sådant att x är en kaka och du tar x (någon gång i framtiden).
> $3''$. $\forall x(Kx \rightarrow P\underline{F}Tdx)$. Det gäller för alla x att om x är en kaka, så är det tillåtet att du tar x (någon gång i framtiden).
> $4''$. $P\forall x(Kx \rightarrow \underline{F}Tdx)$. Det är tillåtet att det gäller för alla x att om x är en kaka, så tar du x (någon gång i framtiden).

Och dessa satser är i ett sådant system inte helt logiskt oberoende av varandra. Tvärt om, man kan notera flera intressanta implikationer. Låt oss sammanfatta dessa i ett teorem.

Teorem 2. **(i)** ($1''$) $\exists x(Kx \land P\underline{F}Tdx)$ och ($2''$) $P\exists x(Kx \land \underline{F}Tdx)$ är logiskt ekvivalenta. **(ii)** ($3''$) $\forall x(Kx \rightarrow P\underline{F}Tdx)$ medför ($1''$) $\exists x(Kx \land P\underline{F}Tdx)$ och ($2''$) $P\exists x(Kx \land \underline{F}Tdx)$ givet att det finns åtminstone ett K. **(iii)** ($4''$) $P\forall x(Kx \rightarrow \underline{F}Tdx)$ medför ($3''$) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. **(iv)** ($4''$) $P\forall x(Kx \rightarrow \underline{F}Tdx)$ medför ($1''$) $\exists x(Kx \land P\underline{F}Tdx)$ och ($2''$) $P\exists x(Kx \land \underline{F}Tdx)$ om det finns åtminstone ett K. **(v)** Varken ($1''$) $\exists x(Kx \land P\underline{F}Tdx)$ eller ($2''$) $P\exists x(Kx \land \underline{F}Tdx)$ medför ($3''$) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. **(vi)** Det är inte fallet att ($3''$) $\forall x(Kx \rightarrow P\underline{F}Tdx)$ medför ($4''$) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. **(vii)** Det är varken fallet att ($1''$) $\exists x(Kx \land P\underline{F}Tdx)$ eller ($2''$) $P\exists x(Kx \land \underline{F}Tdx)$ medför ($4''$) $P\forall x(Kx \rightarrow \underline{F}Tdx)$.

Bevis. Jag skall gå igenom några av dessa punkter. **(i)** Vi visar att $P\exists x(Kx \wedge \underline{F}Tdx)$ och $\exists x(Kx \wedge P\underline{F}Tdx)$ är ekvivalenta genom att först bevisa att $P\exists x(Kx \wedge \underline{F}Tdx)$ medför $\exists x(Kx \wedge P\underline{F}Tdx)$ och sedan att $\exists x(Kx \wedge P\underline{F}Tdx)$ medför $P\exists x(Kx \wedge \underline{F}Tdx)$.

$$
\begin{array}{c}
P\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
\neg\exists x(Kx \wedge P\underline{F}Tdx),\ w_0t_0 \\
\forall x\neg(Kx \wedge P\underline{F}Tdx),\ w_0t_0 \\
sw_0w_1t_0 \\
\exists x(Kx \wedge \underline{F}Tdx),\ w_1t_0 \\
rw_0w_1t_0 \\
Kc \wedge \underline{F}Tdc,\ w_1t_0 \\
Kc,\ w_1t_0 \\
\underline{F}Tdc,\ w_1t_0 \\
\neg(Kc \wedge P\underline{F}Tdc),\ w_0t_0 \\
\swarrow \qquad \searrow \\
\begin{array}{cc}
\neg Kc,\ w_0t_0 & \neg P\underline{F}Tdc,\ w_0t_0 \\
Kc,\ w_0t_0 & O\neg\underline{F}Tdc,\ w_0t_0 \\
* & \neg\underline{F}Tdc,\ w_1t_0 \\
 & *
\end{array}
\end{array}
$$

Följande semantiska tablå bevisar att $\exists x(Kx \wedge P\underline{F}Tdx)$ medför $P\exists x(Kx \wedge \underline{F}Tdx)$.

$$
\begin{array}{c}
\exists x(Kx \wedge P\underline{F}Tdx),\ w_0t_0 \\
\neg P\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
O\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
Kc \wedge P\underline{F}Tdc,\ w_0t_0 \\
Kc,\ w_0t_0 \\
P\underline{F}Tdc,\ w_0t_0 \\
sw_0w_1t_0 \\
\underline{F}Tdc,\ w_1t_0 \\
rw_0w_1t_0 \\
\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_1t_0 \\
\forall x\neg(Kx \wedge \underline{F}Tdx),\ w_1t_0 \\
\neg(Kc \wedge \underline{F}Tdc),\ w_1t_0 \\
\swarrow \qquad \searrow \\
\begin{array}{cc}
\neg Kc,\ w_1t_0 & \neg\underline{F}Tdc,\ w_1t_0 \\
Kc,\ w_1t_0 & * \\
* &
\end{array}
\end{array}
$$

(ii) Låt oss bevisa att $\forall x(Kx \to P\underline{F}Tdx)$ och $\exists xKx$ medför $\exists x(Kx \wedge P\underline{F}Tdx)$. Det är enkelt att se att denna härledning är giltig av rent predikatlogiska skäl. Eftersom $P\exists x(Kx \wedge \underline{F}Tdx)$ är ekvivalent med $\exists x(Kx \wedge P\underline{F}Tdx)$ följer det omedelbart att $\forall x(Kx \to P\underline{F}Tdx)$ och $\exists xKx$ också medför $P\exists x(Kx \wedge \underline{F}Tdx)$. Detta bevis kräver inga särskilda antaganden.

$\exists xKx, w_0t_0$
$\forall x(Kx \to P\underline{F}Tdx), w_0t_0$
$\neg\exists x(Kx \wedge P\underline{F}Tdx), w_0t_0$
$\forall x\neg(Kx \wedge P\underline{F}Tdx), w_0t_0$
Kc, w_0t_0
$Kc \to P\underline{F}Tdc, w_0t_0$
$\neg(Kc \wedge P\underline{F}Tdc), w_0t_0$
↙ ↘
$\neg Kc, w_0t_0$ | $P\underline{F}Tdc, w_0t_0$
* | ↙ ↘
$\neg Kc, w_0t_0$ | $\neg P\underline{F}Tdc, w_0t_0$
* | *

(iii) Följande semantiska tablå bevisar att $P\forall x(Kx \to \underline{F}Tdx)$ medför $\forall x(Kx \to P\underline{F}Tdx)$.

$P\forall x(Kx \to \underline{F}Tdx), w_0t_0$
$\neg\forall x(Kx \to P\underline{F}Tdx), w_0t_0$
$\exists x\neg(Kx \to P\underline{F}Tdx), w_0t_0$
$\neg(Kc \to P\underline{F}Tdc), w_0t_0$
Kc, w_0t_0
$\neg P\underline{F}Tdc, w_0t_0$
$O\neg\underline{F}Tdc, w_0t_0$
$sw_0w_1t_0$
$\forall x(Kx \to \underline{F}Tdx), w_1t_0$
$rw_0w_1t_0$
$Kc \to \underline{F}Tdc, w_1t_0$
↙ ↘
$\neg Kc, w_1t_0$ | $\underline{F}Tdc, w_1t_0$
Kc, w_1t_0 | $\neg\underline{F}Tdc, w_1t_0$
* | *

(iv) Eftersom $\forall x(Kx \to P\underline{F}Tdx)$ och $\exists xKx$ medför både $P\exists x(Kx \wedge \underline{F}Tdx)$ och $\exists x(Kx \wedge P\underline{F}Tdx)$ följer det omedelbart (från steg (iii)) att $P\forall x(Kx \to \underline{F}Tdx)$ och $\exists xKx$ medför både $P\exists x(Kx \wedge \underline{F}Tdx)$ och $\exists x(Kx \wedge P\underline{F}Tdx)$. ■

Det gäller alltså (i dessa system) att ”Du får ta alla kakor” är starkare än ”Du får ta vilken kaka som helst”, och att ”Du får ta vilken kaka som helst” är starkare än ”Du får ta en kaka” (då denna sats tolkas som 1″ eller 2″) och vi antar att det finns åtminstone en kaka.

Vad har nu dessa fakta med fritt val tillåtelser paradoxen att göra och vårt nuvarande problem med den analys jag har föreslagit?

Eftersom ”Du får ta vilken kaka som helst” är starkare än ”Du får ta en kaka” tolkad som 1″ eller 2″ (givet att det finns åtminstone en kaka), är det naturligt att börja med att tolka ”Du får ta en kaka” som ”Du får ta en kaka, vilken som helst”, dvs. ”Du får ta vilken kaka som helst”. Men detta kan dras tillbaka i vissa kontexter. Om någon säger ”Du får ta en kaka, men jag vet inte vilken” är det rimligt att tolka denna sats som 1″ eller 2″.

Likadant tycks det förhålla sig i vårt scenario. Antag att vår nya servitris säger ”Du får ta en tårtbit eller en bakelse (till efterrätt). Men jag vet inte vilket, jag är ny här. Låt mig kolla med köket”. Då tycks det vara rimligt att symbolisera ”Du får ta en tårtbit eller en bakelse på följande sätt $P(\exists x(Tx \wedge \underline{F}Tdx) \vee \exists y(By \wedge \underline{F}Tdy))$ och inte som $\forall x((Tx \vee Bx) \rightarrow P\underline{F}Tdx)$.

Ett potentiellt problem med denna tolkning är dock att servitrisen kanske inte enbart vill säga att det finns en särskild tårtbit eller en särskild bakelse som det är tillåtet att ta, utan att det antingen är tillåtet att du tar vilken tårtbit som helst eller tillåtet att du tar vilken bakelse som helst.

Men om detta är fallet, kan vi helt enkelt tolka ”Du får ta en tårtbit eller en bakelse” som ett elliptiskt uttryck, dvs. som en förkortning av ”Du får ta en tårtbit eller också får du ta en bakelse”, vilket i detta fall säger samma sak som ”Du får ta vilken tårtbit som helst eller också får du ta vilken bakelse som helst”.

Den här invändningen mot vår analys tycks alltså kunna bemötas. Låt oss nu istället undersöka (vad som förefaller vara) en allvarligare invändning.

5. Problem 2: Inget fritt val

”Du får ta vilken kaka som helst” medför att du får ta kaka 1, att du får ta kaka 2, att du får ta kaka 3… osv. Detta utesluter inte att det är obligatoriskt att du tar alla kakor. ”Det är obligatoriskt att du tar alla kakor” medför att det är obligatoriskt att du tar kaka 1, att det är obligatoriskt att du tar kaka 2, att det är obligatoriskt att du tar kaka 3… osv. Men om det är obligatoriskt att du tar alla kakor, så tycks du inte ha något (moraliskt) fritt val. Du är inte fri att välja bland kakorna. Att använda $\forall x(Kx \rightarrow PTdx)$ som en symbolisering av en fritt val tillåtelse tycks därför inte vara helt tillfredsställande. Att det gäller för varje kaka att du får ta den garanterar inte att du är (moraliskt) fri att välja bland kakorna.

Hur skall man ställa sig till detta problem?

6. Lösningsförslag 2

Jag har hittills använt operatorn P för att uttrycka att någonting är tillåtet. En anledning är att detta angreppssätt dominerar i litteraturen om (FVT) paradoxen. Ett annat skäl är att det torde vara lättare att förstå följande lösningsförslag om man först har gått igenom det första något enklare förslaget. (Se också avsnitt 8 för ytterligare en möjlig anledning.) Mitt andra lösningsförslag bygger på att det ofta är naturligare att använda operatorn K (deontisk kontingens) än P i symboliseringen av fritt val tillåtelser. K kan definieras i termer av P på följande sätt: $KA \leftrightarrow (PA \wedge P\neg A)$. KA läses ”Det är deontiskt kontingent att A”, ”A är frivillig”, ”Det är ett fritt val huruvida A eller ej”, ”Det är tillåtet [i den meningen att det är ett fritt val] att A”…

Antag att en lärare säger till sina elever: ”Lektionen är frivillig” eller ”Du får fritt välja om du skall gå på lektionen eller inte”. För att symbolisera denna tillåtelse är det naturligt att använda K. ”Lektionen är frivillig” tycks kunna användas för att säga samma sak som ”Du får fritt välja om du skall gå på lektionen eller inte”, som förefaller säga samma sak som ”Det är tillåtet att du går på lektionen och det är tillåtet att du inte går på lektionen”. Antag att G står för ”Du går på lektionen”. Då symboliseras ”Du får fritt välja om du skall gå på lektionen eller ej” på följande sätt: KG. KG medför $PG \wedge P\neg G$ och $PG \wedge P\neg G$ medför KG. Detta är intuitivt rimligt. Följande argument tycks t.ex. vara giltiga.

> Du får fritt välja om du skall gå på lektionen eller inte. Alltså, det är tillåtet att du går på lektionen och det är tillåtet att du inte går på lektionen.

> Det är tillåtet att du går på lektionen och det är tillåtet att du inte går på lektionen. Alltså, du får fritt välja om du skall gå på lektionen eller inte.

PG är förenlig med OG, dvs. det är tillåtet att du går på lektionen är förenligt med att det är obligatoriskt att du går på lektionen. Att använda P för att symbolisera ”Lektionen är frivillig” eller ”Du får fritt välja om du skall gå på lektionen eller inte” är därför suspekt. Följande yttrande är märkligt: ”Lektionen är frivillig, men ni måste gå på den” och om en lärare sa detta till sina elever skulle de nog ha svårt att förstå vad hon egentligen menade. ”Det är tillåtet att ni går på lektionen… Faktum är att det är obligatoriskt” tycks däremot kunna uttrycka ett påstående som är sant.

Ibland tycks det också vara rimligt att använda K för att symbolisera satser som innehåller ord som ”får”, ”tillåtet” och liknande. Återigen, antag att du är hos din farmor och farfar och att farmor sträcker fram ett fat fyllt med en massa bullar och samtidigt säger ”Du får ta en bulle”. En möjlig symbolisering av denna sats som använder K skulle kunna se ut på följande sätt: $\forall x(Bx \rightarrow K\underline{F}Tdx) = \forall x(Bx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$. Denna formella sats läses ”För varje bulle [på fatet] gäller det att det är tillåtet att du tar den (någon gång i framtiden) och att det är tillåtet att du inte tar den (någon gång i framtiden)” eller ”För varje x gäller det att om x är en bulle [på fatet], så är det tillåtet att du tar x (någon gång i framtiden) och det är tillåtet att du inte tar x (någon gång i framtiden)”. $\forall x(Bx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$ är ekvivalent med $\forall x(Bx \rightarrow P\underline{F}Tdx) \wedge \forall x(Bx \rightarrow P\neg\underline{F}Tdx)$, som läses ”Det gäller för varje x att om x är en bulle så är det tillåtet att du tar x (någon gång i framtiden) och det gäller för alla x att om x är en bulle så är det tillåtet att du inte tar x (någon gång i framtiden)”, dvs. det är tillåtet att du tar vilken bulle som helst och det är tillåtet att du låter bli att ta vilken bulle som helst. Om vi symboliserar ”Du får ta en bulle” på detta sätt och antar att det finns åtminstone en bulle, så utesluter det att det är obligatoriskt att du tar alla bullar. Det här kommer mycket nära vad vi ofta avser med en fritt val tillåtelse tror jag.

Vi visar nu att $\exists xBx$ och $\forall x(Bx \rightarrow P\neg\underline{F}Tdx)$ medför $\neg O\forall x(Bx \rightarrow \underline{F}Tdx)$. Det är lätt att se att det följer att $\exists xBx$ och $\forall x(Bx \rightarrow K\underline{F}Tdx)$ medför $\neg O\forall x(Bx \rightarrow \underline{F}Tdx)$. Med andra ord, om vi antar att det finns åtminstone en bulle och det gäller för varje bulle att du fritt får välja om du skall ta den eller inte, så är det inte obligatoriskt att du tar alla bullar.

$\exists xBx,\ w_0t_0$
$\forall x(Bx \rightarrow P\neg\underline{F}Tdx),\ w_0t_0$
$\neg\neg O\forall x(Bx \rightarrow \underline{F}Tdx),\ w_0t_0$
$O\forall x(Bx \rightarrow \underline{F}Tdx),\ w_0t_0$
$Bc,\ w_0t_0$
$Bc \rightarrow P\neg\underline{F}Tdc,\ w_0t_0$
↙ ↘

$\neg Bc,\ w_0t_0$	$P\neg\underline{F}Tdc,\ w_0t_0$
*	$sw_0w_1t_0$
	$\neg\underline{F}Tdc,\ w_1t_0$
	$rw_0w_1t_0$
	$\forall x(Bx \rightarrow \underline{F}Tdx),\ w_1t_0$
	$Bc \rightarrow \underline{F}Tdc,\ w_1t_0$
	↙ ↘

$\neg Bc,\ w_1t_0$	$\underline{F}Tdc,\ w_1t_0$
$Bc,\ w_1t_0$	*
*	

Betrakta på nytt satserna 1-4. Om vi använder K istället för P för att symbolisera dessa satser kan vi formalisera 1-4 på följande sätt.

$1'''$. $\exists x(Kx \wedge K\underline{F}Tdx)$ [$\Leftrightarrow \exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$]. $1'''$ läses ”Det finns ett x sådant att x är en kaka och det är tillåtet att det någon gång i framtiden kommer att vara fallet att du tar x och det är tillåtet att det inte är fallet att det någon gång i framtiden kommer att vara fallet att du tar x”, eller, förenklat, ”Det finns en kaka sådan att du fritt får välja om du skall ta den eller inte (någon gång i framtiden)”.

$2'''$. $K\exists x(Kx \wedge \underline{F}Tdx)$ [$\Leftrightarrow P\exists x(Kx \wedge \underline{F}Tdx) \wedge P\neg\exists x(Kx \wedge \underline{F}Tdx)$]. $2'''$ läses ”Det är tillåtet att det finns ett x sådant att x är en kaka och det kommer någon gång i framtiden vara fallet att du tar x och det är tillåtet att det inte är fallet att det finns ett x sådant att x är en kaka och det kommer någon gång i framtiden vara fallet att du tar x”. Förenklat kan detta uttryckas ”Du får fritt välja om du skall ta en kaka eller inte (någon gång i framtiden)”.

$3'''$. $\forall x(Kx \rightarrow K\underline{F}Tdx)$ [$\Leftrightarrow \forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$]. $3'''$ läses ”Det gäller för alla x att om x är en kaka, så är det tillåtet att det någon gång i framtiden kommer vara fallet att du tar x och det är tillåtet att det inte är fallet att det någon gång i framtiden kommer vara fallet att du tar x”. $3'''$ kan också, förenklat, läsas på följande sätt ”För varje kaka gäller det att du fritt får välja om du skall ta den eller inte (någon gång i framtiden)”.

$4'''$. $K\forall x(Kx \rightarrow \underline{F}Tdx)$ [$\Leftrightarrow P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge P\neg\forall x(Kx \rightarrow \underline{F}Tdx)$]. $4'''$ tolkas på följande sätt ”Det är tillåtet att det gäller för alla x att om x är en kaka, så kommer det någon gång i framtiden vara fallet att du tar x och det är tillåtet att det inte är fallet att det gäller för alla x att om x är en kaka, så kommer det någon gång i framtiden vara fallet att du tar x”. Förenklat kan detta uttryckas ”Du får fritt välja om du skall ta alla kakor eller inte (någon gång i framtiden)”.

Notera att det inte är fallet att $1'''$ och $2'''$ är logiskt ekvivalenta och att det inte är fallet att $4'''$ är starkare än $3'''$. Däremot finns det flera andra intressanta relationer mellan dessa formler och mellan dessa formler och tidigare symboliseringar av 1-4. Låt oss sammanfatta några av dessa i ett par teorem.

Teorem 3. **(i)** ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$ och $\exists xKx$ medför ($1'''$) $\exists x(Kx \wedge K\underline{F}Tdx)$. **(ii)** ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$, som per definition är ekvivalent med $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$, är logiskt ekvivalent med $\forall x(Kx \rightarrow P\underline{F}Tdx) \wedge \forall x(Kx \rightarrow P\neg\underline{F}Tdx)$. **(iii)** ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$ medför ($3''$) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. **(iv)** ($2'''$) $K\exists x(Kx \wedge \underline{F}Tdx)$ medför ($1'''$) $\exists x(Kx \wedge K\underline{F}Tdx)$. **(v)** ($4''$) $P\forall x(Kx \rightarrow \underline{F}Tdx)$ och $P\neg\exists x(Kx \wedge \underline{F}Tdx)$ medför ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$. **(vi)** $P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge \neg O\exists x(Kx \wedge \underline{F}Tdx)$ medför ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$. **(vii)** $P\forall x(Kx \rightarrow Tdx) \wedge P\forall x(Kx \rightarrow \neg Tdx)$ medför ($3'''$) $\forall x(Kx \rightarrow K\underline{F}Tdx)$.

(viii) $(4''')$ $K\forall x(Kx \rightarrow \underline{F}Tdx)$ och $K\neg\exists x(Kx \wedge \underline{F}Tdx)$ medför $(3''')$ $\forall x(Kx \rightarrow K\underline{F}Tdx)$. **(ix)** $(1''')$ $\exists x(Kx \wedge K\underline{F}Tdx)$ medför $(1'')$ $\exists x(Kx \wedge P\underline{F}Tdx)$.

Bevis. Låt oss bevisa några av dessa punkter. **(i)** $(3''')$ $\forall x(Kx \rightarrow K\underline{F}Tdx)$ och $\exists xKx$ medför $(1''')$ $\exists x(Kx \wedge K\underline{F}Tdx)$. Dvs. följande argument är giltigt. Det gäller för varje kaka att du fritt får välja om du skall ta den eller inte (någon gång i framtiden). Det finns åtminstone en kaka. Alltså finns det en kaka som du fritt får välja om du skall ta eller inte (någon gång i framtiden). Detta argument är giltigt av rent predikatlogiska skäl. Notera att $\forall x(Kx \rightarrow K\underline{F}Tdx)$ per definition är ekvivalent med $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$ och att $\exists x(Kx \wedge K\underline{F}Tdx)$ per definition är ekvivalent med $\exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$. Alltså är också följande argument giltigt. För varje kaka gäller det att det är tillåtet att du tar den (någon gång i framtiden) och att det är tillåtet att du inte tar den (någon gång i framtiden). Alltså finns det en kaka sådan att det är tillåtet att du tar den (någon gång i framtiden) och det är tillåtet att du inte tar den (någon gång i framtiden). Jag har använt mig av dessa fakta i steg 4 och 5 i nedanstående semantiska tablå.

$$
\begin{array}{c}
\exists xKx,\ w_0t_0 \\
\forall x(Kx \rightarrow K\underline{F}Tdx),\ w_0t_0 \\
\neg\exists x(Kx \wedge K\underline{F}Tdx),\ w_0t_0 \\
\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
\neg\exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
\forall x\neg(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
Kc,\ w_0t_0 \\
Kc \rightarrow (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc),\ w_0t_0 \\
\neg(Kc \wedge (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc)),\ w_0t_0 \\
(P\underline{F}Tdc \wedge P\neg\underline{F}Tdc),\ w_0t_0 \\
\swarrow \qquad\qquad\qquad \searrow \\
\begin{array}{cc}
\neg Kc,\ w_0t_0 & \neg(P\underline{F}Tdc \wedge P\neg\underline{F}Tdc),\ w_0t_0 \\
* & *
\end{array}
\end{array}
$$

(ii) $(3''')$ $\forall x(Kx \rightarrow K\underline{F}Tdx)$ är per definition ekvivalent med $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$, och denna sats är logiskt ekvivalent med $\forall x(Kx \rightarrow P\underline{F}Tdx) \wedge \forall x(Kx \rightarrow P\neg\underline{F}Tdx)$, vilket vi kan bevisa med hjälp av vanliga predikatlogiska regler. Detta innebär att ”Det gäller för varje kaka att du fritt får välja om du skall ta den eller inte (någon gång i framtiden)” är ekvivalent med ”Det gäller för varje x att om x är en kaka så är det tillåtet att du tar x (någon gång i framtiden) och det gäller för varje x att om x är en kaka så är det tillåtet att du inte tar x (någon gång i framtiden)”.

(iii) $(3''')$ $\forall x(Kx \rightarrow K\underline{F}Tdx)$ medför $(3'')$ $\forall x(Kx \rightarrow P\underline{F}Tdx)$. Detta följer omedelbart ifrån steg (ii) ovan. Däremot gäller inte det omvända, dvs. det är inte fallet att $(3'')$ medför $(3''')$. $(3''')$ är alltså starkare än $(3'')$.

(iv) $(2''')$ $K\exists x(Kx \wedge \underline{F}Tdx)$ medför $(1''')$ $\exists x(Kx \wedge K\underline{F}Tdx)$. Dvs. följande argument är giltigt. Du får fritt välja om du skall ta en kaka eller inte (någon gång i framtiden). Alltså finns det en kaka som du fritt får välja om du skall ta eller inte (någon gång i framtiden).

Nedan visar vi att det omvända inte gäller (se teorem 4). Dvs. det är inte fallet att $(1''')$ medför $(2''')$. $(2''')$ är alltså starkare än $(1''')$.

Notera att $K\exists x(Kx \wedge \underline{F}Tdx)$ per definition är ekvivalent med $P\exists x(Kx \wedge \underline{F}Tdx) \wedge P\neg\exists x(Kx \wedge \underline{F}Tdx)$, och att $\exists x(Kx \wedge K\underline{F}Tdx)$ per definition är ekvivalent med $\exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$. Jag använder mig av dessa fakta i steg 1, 2 och 3 i nedanstående semantiska tablå.

$$
\begin{array}{c}
P\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
P\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
\neg\exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
\forall x\neg(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
sw_0w_1t_0 \\
\exists x(Kx \wedge \underline{F}Tdx),\ w_1t_0 \\
rw_0w_1t_0 \\
Kc \wedge \underline{F}Tdc,\ w_1t_0 \\
Kc,\ w_1t_0 \\
\underline{F}Tdc,\ w_1t_0 \\
Kc,\ w_0t_0 \\
sw_0w_2t_0 \\
\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_2t_0 \\
rw_0w_2t_0 \\
\forall x\neg(Kx \wedge \underline{F}Tdx),\ w_2t_0 \\
Kc,\ w_2t_0 \\
\neg(Kc \wedge (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc)),\ w_0t_0 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
\begin{array}{c} \neg Kc,\ w_0t_0 \\ * \end{array} &
\begin{array}{c}
\neg(P\underline{F}Tdc \wedge P\neg\underline{F}Tdc),\ w_0t_0 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
\begin{array}{c} \neg P\underline{F}Tdc,\ w_0t_0 \\ O\neg\underline{F}Tdc,\ w_0t_0 \\ \neg\underline{F}Tdc,\ w_1t_0 \\ * \end{array} &
\begin{array}{c} \neg P\neg\underline{F}Tdc,\ w_0t_0 \\ O\neg\neg\underline{F}Tdc,\ w_0t_0 \\ \neg\neg\underline{F}Tdc,\ w_2t_0 \\ \neg(Kc \wedge \underline{F}Tdc),\ w_2t_0 \\ \swarrow \qquad\qquad \searrow \\
\begin{array}{cc} \begin{array}{c} \neg Kc,\ w_2t_0 \\ * \end{array} & \begin{array}{c} \neg\underline{F}Tdc,\ w_2t_0 \\ * \end{array} \end{array}
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
$$

(v) $(4'')$ $P\forall x(Kx \rightarrow \underline{F}Tdx)$ och $P\neg\exists x(Kx \wedge \underline{F}Tdx)$ medför $(3''')$ $\forall x(Kx \rightarrow K\underline{F}Tdx)$. $\forall x(Kx \rightarrow K\underline{F}Tdx)$ är ekvivalent med $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$. Så för att visa (v) räcker det med att visa att premisserna medför

$\forall x(Kx \to (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$. Detta innebär att följande argument är giltigt. Det är tillåtet att du tar alla kakor och det är tillåtet att du inte tar någon kaka alls. Alltså gäller det för varje kaka att du får ta den eller inte. Lite mer exakt, men något osmidigare, kan detta uttryckas på följande sätt. Det är tillåtet att det gäller för varje x att om x är en kaka, så kommer det någon gång i framtiden vara fallet att du tar x. Det är tillåtet att det inte finns något x sådant att x är en kaka och det kommer någon gång i framtiden vara fallet att du tar x. Alltså gäller det för varje x att om x är en kaka, så är det tillåtet att du någon gång i framtiden tar x och det är tillåtet att det inte är fallet att du någon gång i framtiden tar x.

$$
\begin{array}{c}
P\forall x(Kx \to \underline{F}Tdx),\ w_0t_0 \\
P\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_0t_0 \\
\neg\forall x(Kx \to (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
\exists x\neg(Kx \to (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx)),\ w_0t_0 \\
\neg(Kc \to (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc)),\ w_0t_0 \\
Kc,\ w_0t_0 \\
\neg(P\underline{F}Tdc \wedge P\neg\underline{F}Tdc),\ w_0t_0 \\
sw_0w_1t_0 \\
\forall x(Kx \to \underline{F}Tdx),\ w_1t_0 \\
Kc \to \underline{F}Tdc,\ w_1t_0 \\
rw_0w_1t_0 \\
Kc,\ w_1t_0 \\
\underline{F}Tdc,\ w_1t_0 \\
sw_0w_2t_0 \\
\neg\exists x(Kx \wedge \underline{F}Tdx),\ w_2t_0 \\
\forall x\neg(Kx \wedge \underline{F}Tdx),\ w_2t_0 \\
rw_0w_2t_0 \\
Kc,\ w_2t_0 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
\begin{array}{c}
\neg P\underline{F}Tdc,\ w_0t_0 \\
O\neg\underline{F}Tdc,\ w_0t_0 \\
\neg\underline{F}Tdc,\ w_1t_0 \\
*
\end{array}
&
\begin{array}{c}
\neg P\neg\underline{F}Tdc,\ w_0t_0 \\
O\neg\neg\underline{F}Tdc,\ w_0t_0 \\
\neg(Kc \wedge \underline{F}Tdc),\ w_2t_0 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
\begin{array}{c}
\neg Kc,\ w_2t_0 \\
*
\end{array}
&
\begin{array}{c}
\neg\underline{F}Tdc,\ w_2t_0 \\
\neg\neg\underline{F}Tdc,\ w_2t_0 \\
*
\end{array}
\end{array}
\end{array}
\end{array}
\end{array}
$$

$P\neg\exists x(Kx \wedge \underline{F}Tdx)$ är ekvivalent med $\neg O\exists x(Kx \wedge \underline{F}Tdx)$ och med $P\forall x(Kx \to \neg\underline{F}Tdx)$. Därför är också följande två argument giltiga.

(vi) $P\forall x(Kx \to \underline{F}Tdx) \wedge \neg O\exists x(Kx \wedge \underline{F}Tdx)$ medför $\forall x(Kx \to K\underline{F}Tdx)$. Intuitivt, och förenklat, säger detta argument följande. Det är tillåtet att du tar alla kakor, men du behöver inte (måste inte) ta någon kaka alls. Alltså gäller det för varje kaka att du fritt får välja om du skall ta den eller inte.

(vii) $P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge P\forall x(Kx \rightarrow \neg\underline{F}Tdx)$ medför $\forall x(Kx \rightarrow K\underline{F}Tdx)$. Intuitivt, och förenklat, säger detta argument följande. Om det är tillåtet att du tar alla kakor och det är tillåtet att det gäller för varje kaka att du inte tar den, så gäller det för varje kaka att det är tillåtet att du tar den och att det är tillåtet att du inte tar den.

(viii) (4‴) $K\forall x(Kx \rightarrow \underline{F}Tdx)$ och $K\neg\exists x(Kx \wedge \underline{F}Tdx)$ medför (3‴) $\forall x(Kx \rightarrow K\underline{F}Tdx)$. Något förenklat säger detta argument följande. Du får fritt välja om du skall ta alla kakor eller inte och du får fritt välja om du inte skall ta någon kaka alls eller inte. Alltså gäller det för varje kaka att du fritt får välja att ta den eller inte. $K\forall x(Kx \rightarrow \underline{F}Tdx)$ är logiskt ekvivalent med $P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge P\neg\forall x(Kx \rightarrow \underline{F}Tdx)$ och $K\neg\exists x(Kx \wedge \underline{F}Tdx)$ med $P\neg\exists x(Kx \wedge \underline{F}Tdx) \wedge P\neg\neg\exists x(Kx \wedge \underline{F}Tdx)$. Eftersom vi redan har bevisat steg (v) ovan är det lätt att se att detta argument är giltigt.

(ix) (1‴) $\exists x(Kx \wedge K\underline{F}Tdx)$ medför (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$. Intuitivt, och förenklat, säger detta argument. Det finns en kaka som du fritt får välja om du skall ta eller inte. Alltså finns det en kaka som du får ta. Eftersom $\exists x(Kx \wedge K\underline{F}Tdx)$ per definition är ekvivalent med $\exists x(Kx \wedge (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$ är det lätt att se att (1‴) medför (1″) av rent predikatlogiska skäl. Däremot medför (1″) inte (1‴). (1‴) är alltså starkare än (1″). ■

Teorem 4. **(i)** Det är inte fallet att (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$ medför (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. **(ii)** Det är inte fallet att (2″) $P\exists x(Kx \wedge \underline{F}Tdx)$ medför (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. **(iii)** Det är inte fallet att (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. **(iv)** Det är inte fallet att (2″) $P\exists x(Kx \wedge \underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. **(v)** Det är inte fallet att (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. **(vi)** Det är inte fallet att (3‴) $\forall x(Kx \rightarrow K\underline{F}Tdx)$ medför (4‴) $K\forall x(Kx \rightarrow \underline{F}Tdx)$. **(vii)** Det är inte fallet att (1‴) $\exists x(Kx \wedge K\underline{F}Tdx)$ medför (2‴) $\underline{K}\exists x(Kx \wedge \underline{F}Tdx)$. **(viii)** Det är inte fallet att (4‴) $K\forall x(Kx \rightarrow \underline{F}Tdx)$ medför (3‴) $\forall x(Kx \rightarrow K\underline{F}Tdx)$.

Bevis. **(i)** Det är inte fallet att (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$ medför (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. Följande modell visar detta. W är mängden av alla möjliga världar, T är mängden av alla tidpunkter, D är domänen av entiteter vi kvantifierar över, $Sww't$ innebär att w' är deontiskt tillgänglig från w vid t, $Rww't$ att w' är aletiskt tillgänglig från w vid t och $t < t'$ att t inträffar före t' (t' efter t).[6] $W = \{w_0, w_1\}$, $T = \{t_0, t_1\}$, $D = \{c, d, f\}$, $t_0 < t_1$. $Sw_0w_1t_0$, $Sw_1w_1t_0$, $Sw_0w_0t_1$, $Sw_1w_1t_1$. I t_0 är alla möjliga världar aletiskt tillgängliga från alla möjliga världar. $Rw_0w_0t_1$, $Rw_1w_1t_1$. Kc och Kf är sanna i w_0 vid t_0 och i w_1 vid t_0, Tdc

6 För mer information om den typ av semantik jag använder här, se Rönnedal (2012), (2012b) och (2014).

är sann i w_1 vid t_1 och Tdf är falsk i w_1 vid t_1.[7] Sanningsvärdena hos några centrala satser kan nu räknas ut på följande sätt. w_0t_0: Kc, Kf, $P\underline{F}Tdc$, $\neg P\underline{F}Tdf$, $Kc \wedge P\underline{F}Tdc$, $\neg(Kf \rightarrow P\underline{F}Tdf)$, $\exists x(Kx \wedge P\underline{F}Tdx)$, $\neg\forall x(Kx \rightarrow P\underline{F}Tdx)$. w_1t_0: $\underline{F}Tdc$, $\neg\underline{F}Tdf$. w_1t_1: Tdc, ¬Tdf. Låt oss i detalj gå igenom hur denna modell bevisar vår slutsats för att belysa detta sätt att resonera. Eftersom Tdc är sann i w_1 vid t_1 och $t_0 < t_1$, så är $\underline{F}Tdc$ sann i w_1 vid t_0. Det följer att $P\underline{F}Tdc$ är sann i w_0 vid t_0, eftersom $Sw_0w_1t_0$. Tdf är inte sann i w_1 vid t_1. Eftersom t_1 är den enda tidpunkt som inträffar efter t_0 följer det att $\neg\underline{F}Tdf$ är sann i w_1 vid t_0. Alltså är $\neg P\underline{F}Tdf$ sann i w_0 vid t_0, för w_1 är den enda värld som är deontiskt tillgänglig från w_0 vid t_0. Således är $Kc \wedge P\underline{F}Tdc$ sann i w_0 vid t_0. För både Kc och $P\underline{F}Tdc$ är sanna i denna värld vid denna tidpunkt. Det följer att $\exists x(Kx \wedge P\underline{F}Tdx)$ är sann i w_0 vid t_0. Eftersom Kf är sann i w_0 vid t_0 och $P\underline{F}Tdf$ är falsk i denna värld vid denna tidpunkt följer det att $Kf \rightarrow P\underline{F}Tdf$ är falsk i w_0 vid t_0. Således kan vi sluta oss till att $\forall x(Kx \rightarrow P\underline{F}Tdx)$ är falsk i w_0 vid t_0. I denna modell är alltså $\exists x(Kx \wedge P\underline{F}Tdx)$ sann i w_0 vid t_0 och $\forall x(Kx \rightarrow P\underline{F}Tdx)$ falsk i w_0 vid t_0. Det följer att det inte är fallet att (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$ medför (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$.

(ii) Det är inte fallet att (2″) $P\exists x(Kx \wedge \underline{F}Tdx)$ medför (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$. Eftersom $\exists x(Kx \wedge P\underline{F}Tdx)$ inte medför $\forall x(Kx \rightarrow PFTdx)$ och $\exists x(Kx \wedge P\underline{F}Tdx)$ är logisk ekvivalent med $P\exists x(Kx \wedge \underline{F}Tdx)$, så följer inte $\forall x(Kx \rightarrow PFTdx)$ ur $P\exists x(Kx \wedge \underline{F}Tdx)$. Vi kan använda samma modell som i (i) för att visa detta. Då kan vi räkna ut sanningsvärdena hos några centrala satser på följande sätt. w_0t_0: Kc, Kf, $P\exists x(Kx \wedge \underline{F}Tdx)$, $\neg P\underline{F}Tdf$, $\neg(Kf \rightarrow P\underline{F}Tdf)$, $\neg\forall x(Kx \rightarrow P\underline{F}Tdx)$. w_1t_0: Kc, Kf, $\underline{F}Tdc$, $\neg\underline{F}Tdf$, $Kc \wedge \underline{F}Tdc$, $\exists x(Kx \wedge \underline{F}Tdx)$. w_1t_1: Tdc, ¬Tdf.

(iii) Det är inte fallet att (1″) $\exists x(Kx \wedge P\underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. Eftersom $\exists x(Kx \wedge P\underline{F}Tdx)$ inte medför $\forall x(Kx \rightarrow PFTdx)$ och $P\forall x(Kx \rightarrow \underline{F}Tdx)$ medför $\forall x(Kx \rightarrow P\underline{F}Tdx)$, så medför $\exists x(Kx \wedge P\underline{F}Tdx)$ inte $P\forall x(Kx \rightarrow \underline{F}Tdx)$. Vi kan använda samma modell som i (i) för att visa detta.

(iv) Det är inte fallet att (2″) $P\exists x(Kx \wedge \underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. Eftersom $\exists x(Kx \wedge P\underline{F}Tdx)$ inte medför $P\forall x(Kx \rightarrow \underline{F}Tdx)$ och $\exists x(Kx \wedge P\underline{F}Tdx)$ är logiskt ekvivalent med $P\exists x(Kx \wedge \underline{F}Tdx)$, så följer inte $P\forall x(Kx \rightarrow \underline{F}Tdx)$ ur $P\exists x(Kx \wedge \underline{F}Tdx)$. Vi kan använda samma modell som i (i) för att visa detta.

(v) Det är inte fallet att (3″) $\forall x(Kx \rightarrow P\underline{F}Tdx)$ medför (4″) $P\forall x(Kx \rightarrow \underline{F}Tdx)$. Följande modell visar detta. $W = \{w_0, w_1, w_2\}$, $T = \{t_0, t_1\}$, $D = \{c, d, f\}$, $t_0 < t_1$. $Sw_0w_1t_0$, $Sw_0w_2t_0$, $Sw_1w_2t_0$, $Sw_2w_1t_0$, $Sw_1w_1t_0$, $Sw_2w_2t_0$. I t_0 är alla möjliga världar aletiskt tillgängliga från varandra. I t_1 är alla möjliga världar

[7] Övriga sanningsvärden hos satser i olika värld-tidpunkt par är i sammanhanget ointressanta. Man kan anta att de är sanna eller falska, vilket som. Detta gäller även övriga motexempel nedan.

aletiskt och deontiskt tillgängliga från sig själva men inte från någon annan möjlig värld. Vid t_0 är Kc och Kf sanna i varje möjlig värld. Tdc är sann i w_1 vid t_1 och Tdf är falsk. Tdf är sann i w_2 vid t_1 och Tdc falsk. Då kan vi räkna ut följande sanningsvärden hos några centrala satser. w_0t_0: Kc, Kf, $\neg P\forall x(Fx \rightarrow \underline{F}Tdx)$, $P\underline{F}Tdc$, $P\underline{F}Tdf$, $Kc \rightarrow P\underline{F}Tdc$, $Kf \rightarrow P\underline{F}Tdf$, $\forall x(Kx \rightarrow P\underline{F}Tdx)$. w_1t_0: Kc, Kf, $\underline{F}Tdc$, $\neg\underline{F}Tdf$, $\neg(Kf \rightarrow \underline{F}Tdf)$ $\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. w_1t_1: Tdc, $\neg Tdf$. w_2t_0: Kc, Kf, $\underline{F}Tdf$, $\neg\underline{F}Tdc$, $\neg(Kc \rightarrow \underline{F}Tdc)$, $\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. w_2t_1: $\neg Tdc$, Tdf.

(vi) Det är inte fallet att (3‴) $\forall x(Kx \rightarrow K\underline{F}Tdx)$ medför (4‴) $K\forall x(Kx \rightarrow \underline{F}Tdx)$. Detta är ekvivalent med att det inte är fallet att $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$ medför $P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge P\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. Vi kan använda samma modell som i (v) för att visa detta. Här följer de relevanta sanningsvärdena. w_0t_0: Kc, Kf, $\neg P\forall x(Fx \rightarrow \underline{F}Tdx)$, $P\underline{F}Tdc$, $P\underline{F}Tdf$, $P\neg\underline{F}Tdc$, $P\neg\underline{F}Tdf$, $P\underline{F}Tdc \wedge P\neg\underline{F}Tdc$, $P\underline{F}Tdf \wedge P\neg\underline{F}Tdf$, $Kc \rightarrow (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc)$, $Kf \rightarrow (P\underline{F}Tdf \wedge P\neg\underline{F}Tdf)$, $\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$, $\forall x(Kx \rightarrow K\underline{F}Tdx)$, $\neg(P\forall x(Kx \rightarrow \underline{F}Tdx) \wedge P\neg\forall x(Kx \rightarrow \underline{F}Tdx))$, $\neg K\forall x(Kx \rightarrow \underline{F}Tdx)$. w_1t_0: Kc, Kf, $\underline{F}Tdc$, $\neg\underline{F}Tdf$, $\neg(Kf \rightarrow \underline{F}Tdf)$ $\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. w_1t_1: Tdc, $\neg Tdf$. w_2t_0: Kc, Kf, $\underline{F}Tdf$, $\neg\underline{F}Tdc$, $\neg(Kc \rightarrow \underline{F}Tdc)$, $\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. w_2t_1: $\neg Tdc$, Tdf.

(vii) Det är inte fallet att (1‴) $\exists x(Kx \wedge K\underline{F}Tdx)$ medför (2‴) $K\exists x(Kx \wedge \underline{F}Tdx)$. Vi kan använda samma modell som i (v). Här följer de relevanta sanningsvärdena. w_0t_0: Kc, Kf, $\neg P\neg\exists x(Kx \wedge \underline{F}Tdx)$, $\neg(P\exists x(Kx \wedge \underline{F}Tdx) \wedge P\neg\exists x(Kx \wedge \underline{F}Tdx))$, $\neg K\exists x(Kx \wedge \underline{F}Tdx)$, $P\underline{F}Tdc$, $P\neg\underline{F}Tdc$, $K\underline{F}Tdc$, $Kc \wedge K\underline{F}Tdc$, $\exists x(Kx \wedge K\underline{F}Tdx)$. w_1t_0: Kc, Kf, $\underline{F}Tdc$, $\neg\underline{F}Tdf$, $Kc \wedge \underline{F}Tdc$, $\exists x(Kx \wedge \underline{F}Tdx)$. w_1t_1: Tdc, $\neg Tdf$. w_2t_0: Kc, Kf, $\underline{F}Tdf$, $\neg\underline{F}Tdc$, $Kf \wedge \underline{F}Tdf$, $\exists x(Kx \wedge \underline{F}Tdx)$. w_2t_1: $\neg Tdc$, Tdf.

(viii) Det är inte fallet att (4‴) $K\forall x(Kx \rightarrow \underline{F}Tdx)$ medför (3‴) $\forall x(Kx \rightarrow K\underline{F}Tdx)$. För att visa detta kan vi använda samma modell som i (v) med en modifikation. Vi låter Tdc vara sann i w_2 vid t_1 istället för falsk. Då har vi följande relevanta sanningsvärden. w_0t_0: Kc, $P\forall x(Kx \rightarrow \underline{F}Tdx)$, $P\neg\forall x(Kx \rightarrow \underline{F}Tdx)$, $K\forall x(Kx \rightarrow \underline{F}Tdx)$, $\neg P\neg\underline{F}Tdc$, $\neg(P\underline{F}Tdc \wedge P\neg\underline{F}Tdc)$, $\neg(Kc \rightarrow (P\underline{F}Tdc \wedge P\neg\underline{F}Tdc))$, $\neg\forall x(Kx \rightarrow (P\underline{F}Tdx \wedge P\neg\underline{F}Tdx))$, $\neg\forall x(Kx \rightarrow K\underline{F}Tdx)$. w_1t_0: Kc, Kf, $\underline{F}Tdc$, $\neg\underline{F}Tdf$, $\neg(Kf \rightarrow \underline{F}Tdf)$, $\neg\forall x(Kx \rightarrow \underline{F}Tdx)$. w_1t_1: Tdc, $\neg Tdf$. w_2t_0: Kc, Kf, $\underline{F}Tdc$, $\underline{F}Tdf$, $Kc \rightarrow \underline{F}Tdc$, $Kf \rightarrow \underline{F}Tdf$, $\forall x(Kx \rightarrow \underline{F}Tdx)$. w_2t_1: Tdc, Tdf. ■

Jag har föreslagit att fritt val tillåtelser ofta kan symboliseras med en sats som har samma logiska form som (3‴), dvs. $\forall x(Kx \rightarrow K\underline{F}Tdx)$, eller någon variant av denna, t.ex. $\forall x((Bx \vee Mx) \rightarrow K\underline{F}Tdx)$. I ljuset av ovanstående teorem förefaller det vara rimligt. (4‴), $K\forall x(Kx \rightarrow \underline{F}Tdx)$, är för stark för att vara en bra symbolisering av ”Du får ta en kaka”, bl.a. eftersom den medför att du får ta alla kakor. (2‴), $K\exists x(Kx \wedge \underline{F}Tdx)$, tycks också vara för stark i den meningen att den medför $P\neg\exists x(Kx \wedge \underline{F}Tdx)$. Enligt $P\neg\exists x(Kx \wedge \underline{F}Tdx)$ är det tillåtet att du inte tar någon kaka alls. Men fritt val tillåtelser kan ofta kombineras med en disjunktiv plikt. Så, ”Du får ta en kaka” utesluter inte ”Du måste ta åtminstone en kaka”. (Se vidare avsnitt 7 nedan.) Samtidigt tycks $(2''')$ också vara för svag eftersom den endast medför att det finns en specifik kaka som du får ta, inte att du får ta vilken kaka som helst. $(1''')$ $\exists x(Kx \wedge K\underline{F}Tdx)$ kan i vissa sammanhang vara en bra översättning av ”Du får ta en kaka”, men denna formalisering hjälper oss inte att förklara att argument av samma typ som argument 1 tycks vara giltiga. Då återstår $(3''')$. Vi har också sett att $(3''')$ är starkare än $(3'')$ $\forall x(Kx \rightarrow P\underline{F}Tdx)$. $(3''')$ medför $(3'')$ men inte tvärt om. Enligt en plausibel pragmatisk hypotes bör vi tolka varje (yttrande av en) sats på ett sådant sätt att den (det) uttrycker ett så starkt påstående som möjligt. Det kan därför vara rimligt att börja med att tolka satsen ”Du får ta en kaka” som $(3''')$ snarare än som $(3'')$. Om det finns något i kontexten som gör detta orimligt, kan vi istället använda $(3'')$ eller de ännu svagare $(1''')$ eller $(1'')$.

7. Fritt val tillåtelser och disjunktiva plikter

Antag att du är en elev på universitetet och att din lärare i filosofi säger följande: ”Du får skriva en uppsats om Hume eller om Kant. Men du måste skriva en uppsats antingen om Hume eller om Kant.” I denna situation tycks det vara rimligt att tolka den första satsen som en fritt val tillåtelse och den andra satsen som en disjunktiv plikt. Vi tänker oss att ”Du får skriva en uppsats om Hume eller om Kant” säger samma sak som ”Du får fritt välja om du skall skriva en uppsats om Hume eller om Kant”. Om vi symboliserar ”Du får skriva en uppsats om Hume eller om Kant” på det sätt som jag har föreslagit $\forall x((x = h \vee x = k) \rightarrow K\underline{F}Sdx)$, så kan vi härleda följande satser: $P\underline{F}Sdh$, $P\neg\underline{F}Sdh$, $P\underline{F}Sdk$, $P\neg\underline{F}Sdk$. Sxy läses ”x skriver en uppsats om y”, h refererar till Hume och k till Kant. Dvs. vi kan härleda att det är tillåtet att du skriver en uppsats om Hume, att det är tillåtet att det inte är fallet att du skriver en uppsats om Hume, att det är tillåtet att du skriver en uppsats om Kant, och att det är tillåtet att det inte är fallet att du skriver en uppsats om Kant. Satsen ”Du måste skriva en uppsats antingen om Hume eller om Kant” tycks kunna symboliseras på följande sätt: $O(\underline{F}Sdh \vee \underline{F}Sdk)$ ($O(\underline{F}Sdh \vee \underline{F}Sdk)$ är logiskt ekvivalent med $O\underline{F}(Sdh \vee Sdk)$). Frågan är nu om detta är konsistent? Följande modell visar att $\{P\underline{F}Sdh, P\neg\underline{F}Sdh, P\underline{F}Sdk, P\neg\underline{F}Sdk, O(\underline{F}Sdh \vee \underline{F}Sdk)\}$ är konsistent. $W = \{w_0, w_1, w_2, w_3, w_4\}$, $T = \{t_0, t_1, t_2, t_3, t_4\}$, $D = \{d, h, k\}$. $t_0 < t_1$, $t_0 < t_2$, $t_0 < t_3$, $t_0 < t_4$. $Sw_0w_1t_0$, $Sw_0w_2t_0$, $Sw_0w_3t_0$, $Sw_0w_4t_0$. Sdh är sann i w_1 vid t_1. Sdk är sann och Sdh falsk i w_2 vid t_2. Sdk är sann i w_3 vid t_3. Sdh är sann och Sdk falsk i w_4 vid t_4. Mer allmänt gäller det att $\{KA, KB, O(A \vee B)\}$ är konsistent; det är alltså inte fallet att $KA \wedge KB$ medför att $\neg O(A \vee B)$. Så enligt den här analysen är det fullt möjligt att det finns situationer då vi fritt får välja mellan A och B, men måste välja A eller B. Det här är ett positivt besked eftersom fritt val tillåtelser ofta tycks kombineras med disjunktiva plikter. Vårt exempel är bara ett bland i princip oändligt många andra liknande fall.

8. Problem 1: Pragmatiska implikationer trots allt?

Låt oss avsluta med att ta upp två av de intressantaste möjliga invändningar mot de förslag jag har presenterat i den här uppsatsen som jag kan komma på. Enligt den första invändningen är den typ av ”slutledningar” jag har diskuterat, när allt kommer omkring, inte semantiskt giltiga argument, utan ett slags pragmatiska implikationer. $O(A \lor B)$ medför $P(A \lor B)$ i SDL men inte $PA \land PB$. Om någon (som har kunskap om normerna i en viss situation) hävdar att det är obligatoriskt att A eller B, är det emellertid rimligt att pragmatiskt sluta sig till att det är tillåtet att A och att det är tillåtet att B. För OA medför $O(A \lor B)$ och OB medför $O(A \lor B)$. Detta innebär att om vår person visste att det var obligatoriskt att A (eller obligatoriskt att B), så skulle hon hävda att det är obligatoriskt att A (eller att det är obligatoriskt att B) istället för att det är obligatoriskt att A eller B. För ett sådant påstående är mer informativt. Att hävda det svagare påståendet kan vara vilseledande. Eftersom personen i fråga har kunskap om normerna i situationen kan vi sluta oss till att hon vet att det inte är obligatoriskt att A och att hon vet att det inte är obligatoriskt att B. Härav följer det att det inte är obligatoriskt att A och att det inte är obligatoriskt att B. Men $O(A \lor B)$ och $\neg OA$ och $\neg OB$ medför $PA \land PB$ i SDL, vilket bevisas av följande semantiska tablå.

$$
\begin{array}{c}
O(A \lor B), 0 \\
\neg OA, 0 \\
\neg OB, 0 \\
\neg(PA \land PB), 0 \\
P\neg A, 0 \\
P\neg B, 0 \\
0s1 \\
\neg A, 1 \\
0s2 \\
\neg B, 2 \\
\swarrow \qquad\qquad \searrow \\
\begin{array}{cc}
\begin{array}{c}
\neg PA, 0 \\
O\neg A, 0 \\
\neg A, 2 \\
A \lor B, 2 \\
\swarrow \quad \searrow \\
\begin{array}{cc} A, 2 & B, 2 \\ * & * \end{array}
\end{array}
&
\begin{array}{c}
\neg PB, 0 \\
O\neg B, 0 \\
\neg B, 1 \\
A \lor B, 1 \\
\swarrow \quad \searrow \\
\begin{array}{cc} A, 1 & B, 1 \\ * & * \end{array}
\end{array}
\end{array}
\end{array}
$$

Så om din lärare hävdar att det är obligatoriskt att du skriver en uppsats om Hume eller om Kant, så kan vi pragmatiskt sluta oss till att det är tillåtet att du skriver en uppsats om Hume och att det är tillåtet att du skriver en uppsats om Kant. Detta innebär inte att påståendet att det är obligatoriskt att du skriver en uppsats om Hume eller om Kant *medför* att det är tillåtet att du skriver en uppsats om Hume och att det är tillåtet att du skriver en uppsats om Kant.

På liknande sätt förhåller det sig med tillåtelser. PA medför inte P¬A i SDL, men om någon (som har kunskap om de relevanta normerna i en situation) hävder att det är tillåtet att A, så är det rimligt att pragmatiskt sluta sig till att det också är tillåtet att inte A. För OA medför PA i SDL. Och om vår person visste att det var obligatoriskt att A, skulle hon hävda detta istället för att det är tillåtet att A. För påståendet att det är obligatoriskt att A är starkare och därför mer informativt än påståendet att det är tillåtet att A. (Om en kurskamrat till dig vet om att du *måste* skriva en uppsats om Kant, så skulle vi anse att det var vilseledande av henne om hon sa att du *får* skriva en uppsats om Kant, även om hennes påstående inte är falskt och faktiskt följer ur påståendet att du *måste* skriva en uppsats om Kant.) Eftersom personen ifråga har kunskap om de relevanta normerna i situationen kan vi sluta oss till att hon vet att det inte är obligatoriskt att A. Från detta följer det att det inte är obligatoriskt att A. Men ¬OA är logiskt ekvivalent med P¬A i SDL. Alltså kan vi sluta oss till att det är tillåtet att inte A. Så om din farmor säger att du får ta en kaka, så kan vi pragmatiskt sluta oss till att det också är OK om du inte tar en kaka. Detta innebär inte att påståendet att det är tillåtet att du tar en kaka *medför* att det är tillåtet att du inte tar en kaka.

Man skulle alltså kunna hävda att ”Du får ta vilken kaka som helst” pragmatiskt implicerar att du inte måste ta alla kakor men att detta påstående inte följer semantiskt ifrån denna sats. För ”Du får ta vilken kaka som helst” medför ”Du får ta kaka 1 och du får ta kaka 2 och du får…”. Och om någon (med kunskap om normerna i den aktuella situationen) hävdar att du får ta kaka 1 och att du får ta kaka 2 och att du får…, så implicerar detta pragmatiskt att det är tillåtet att du inte tar kaka 1, att det är tillåtet att du inte tar kaka 2 osv. Dvs. det gäller för varje kaka att det är tillåtet att du inte tar den. Och från detta tycks det följa pragmatiskt att du inte måste ta alla kakor. För påståendet att det gäller för varje kaka att det är tillåtet att du inte tar den medför påståendet att det inte är obligatoriskt att du tar alla kakor om vi antar att det finns åtminstone en kaka. Så, lösningsförlag 2 är kanske inte rimligt när allt kommer omkring. Och bör man inte på ett liknande pragmatiskt sätt försöka förklara intuitionen att argument 1 tycks vara giltigt?

Jag är inte helt övertygad om att dessa ”slutsatser” endast är pragmatiska. Om din lärare hävdar att du fritt får välja om du skall skriva en uppsats om Hume eller Kant men att du måste skriva en uppsats om Kant, så tycks hon inte endast påstå något (pragmatiskt) konstigt, utan hon tycks hävda något som är inkonsistent. ”Du får fritt välja om du skall skriva en uppsats om Hume eller Kant” tycks medföra att det är tillåtet att du inte skriver en uppsats om Hume och att det är tillåtet att du inte skriver en uppsats om Kant. Det förefaller åtminstone vara uppenbart att en person som hävdar en sådan sats sannolikt också *menar* att detta är tillåtet. Men kanske finns det här en skillnad mellan ”Du får fritt välja om du skall skriva en uppsats om Hume eller Kant” och ”Du får skriva en uppsats om Hume eller Kant”, kanske bestämmer inte talarens intentioner, åsikter och andra mentala tillstånd meningen hos dessa satser, kanske är intuitionen att satsen ”Du får fritt välja om du skall skriva en uppsats om Hume eller om Kant, men du måste skriva en uppsats om Kant” är inkonsistent inte riktig. Låt oss anta det. Då kan vi återgå till lösningsförslag 1, och hävda att övriga implikationer endast är pragmatiska och inte semantiska.

De pragmatiska implikationer som jag har nämnt ovan förefaller vara rimliga. Från detta följer inte att FVT paradoxen kan lösas helt och hållet med hjälp av pragmatiska principer. Jag kan inte i detalj här diskutera alla pragmatiska lösningar på FVT paradoxen som har föreslagits i litteraturen. Men låt mig ta upp ett generellt problem som talar för att det är rimligt att vara skeptisk till lösningar av detta slag. PA medför $P(A \vee B)$ och PB medför $P(A \vee B)$ i SDL. Därför är det, med tanke på vanliga pragmatiska principer, rimligt att anta att en person (som har kunskap om de relevanta normerna i en situation och) som hävdar att det är tillåtet att A eller B både vet att det inte är tillåtet att A och vet att det inte är tillåtet att B (givet att $P(A \vee B)$ är en korrekt symbolisering av ”Det är tillåtet att A eller B”). För om hon visste att det var tillåtet att A (eller om hon visste att det var tillåtet att B), så skulle hon hävda det starkare påståendet att det är tillåtet att A (tillåtet att B) istället för det svagare påståendet att det är tillåtet att A eller B. Att hävda ett svagare påstående när man vet att ett starkare påstående är sant kan vara vilseledande. Alltså kan vi pragmatiskt sluta oss till att det inte är tillåtet att A och att det inte är tillåtet att B. Dessa pragmatiska resonemang står därför i direkt motsats till våra intuitioner om vad som följer ur vad. ”Det är tillåtet att A eller B”, tycks ju tvärt emot de pragmatiska resonemangen medföra att det är tillåtet att A och att det är tillåtet att B. Och inte nog med det. $P(A \vee B)$ medför $PA \vee PB$ i SDL. Så, våra pragmatiska resonemang tycks leda till en direkt motsägelse.

Pragmatiska teorier och lösningar på FVT paradoxen har under senare år blivit alltmer sofistikerade, och det finns anledning att tro att de kommer att fortsätta att utvecklas framöver. Det är emellertid tveksamt om det i dagsläget finns någon helt tillfredsställande pragmatisk lösning på denna paradox.

9. Problem 2: Kvantifiering över handlingar?

Låt oss ta upp en sista invändning mot de lösningsförslag som har presenterats i den här uppsatsen. Betrakta följande argument.

> Argument 5
> Du får gå till en strand eller (stanna inne och) spela ett dataspel.
> Alltså får du gå till en strand och du får (stanna inne och) spela ett dataspel.

Slutsatsen tycks följa ur premissen. Enligt en naturlig tolkning av detta argument, i linje med tidigare analyser, säger det samma sak som följande härledning: Du får gå till vilken strand som helst eller spela vilket dataspel som helst. Alltså får du gå till vilken strand som helst och du får spela vilket dataspel som helst. Men det tycks inte gå att symbolisera detta resonemang på samma sätt som tidigare argument. Antag att vi försöker med en formalisering som innehåller Sx (är en strand) och Dx (är ett dataspel) och börjar $\forall x((Sx \vee Dx) \rightarrow P\ldots)$. Men vad skall vi då fylla i punkterna med? De olika disjunkterna talar om olika typer av handlingar eller förhållanden som bör symboliseras med olika predikat, t.ex. Gxy (x går till y) och Sxy (x spelar y). Det går att hitta otaliga exempel av detta slag. Man kan också fråga sig vad det egentligen är som är tillåtet här. Är det tillåtet att gå till vilken strand som helst och tillåtet att spela vilket dataspel som helst, eller är det själva aktiviteterna att gå till en strand och spela ett dataspel som är tillåtna?

Flera andra argument väcker liknande frågor. Betrakta på nytt slutsatsen i argument 3: ”Alltså får du gå till stranden och du får gå på bio”. Jag föreslog tidigare att denna sats säger samma sak som satsen ”Du får gå till vilken strand som helst och du får gå till vilken biograf som helst”. Men man skulle kunna hävda att denna sats egentligen handlar om aktiviteten att gå till stranden (och bada) och aktiviteten att gå på bio (och se en film). Enligt denna tolkning säger slutsatsen att det är tillåtet att utföra aktiviteten (handlingen) att gå till stranden (och bada) och att det är tillåtet att utföra aktiviteten (handlingen) att gå på bio (och se en film). Det här antyder en möjlig lösning på det aktuella problemet.

Enligt denna lösning kvantifierar vi inte över platser eller spel i argument 5, utan över handlingar. Låt S och D representera egenskaper hos handlingar, dvs. Sx står för ”x är ett gående till en strand”, Dx för ”x är ett spelande av ett dataspel”, d refererar till dig, och Uxy står för ”x utför y”. Då kan argument 5 symboliseras på följande sätt.

> $\forall x((Sx \vee Dx) \rightarrow PUdx)$.
> Alltså: $\forall x(Sx \rightarrow PUdx) \wedge \forall x(Dx \rightarrow PUdx)$

Premissen läses ”Det gäller för alla x att om x är ett gående till en strand eller x är ett spelande av ett dataspel, så är det tillåtet att du utför x”. Slutsatsen läses: ”Det gäller för alla x att om x är ett gående till en strand så är det tillåtet att du utför x och det gäller för alla x att om x är ett spelande av ett dataspel så är det tillåtet att du utför x”. Detta argument är giltigt i SDL kombinerat med predikatlogik. I ett kvantifierat temporalt aletiskt-deontiskt system med determinerat förflutet (och nu) kan vi lägga till operatorn $\underline{F}$ och symbolisera argumentet på följande sätt.

$\forall x((Sx \lor Dx) \to P\underline{F}Udx)$
Alltså: $\forall x(Sx \to P\underline{F}Udx) \land \forall x(Dx \to P\underline{F}Udx)$

Även i detta fall följer slutsatsen ur premissen. Övriga argument av samma typ tycks kunna analyseras på liknande sätt. Om det är riktigt, kan den här invändningen bemötas.

Ett potentiellt problem med denna lösning är att man kan fråga sig om det någonsin är fallet att premissen i ett argument av detta slag är sann. Betyder t.ex. inte premissen i argument 5, enligt denna läsning, att alla sätt att gå till stranden är tillåtna och alla sätt att spela dataspel är tillåtna? Och är det rimligt? Finns det inte vissa handlingar av dessa typer som det inte är tillåtet att utföra? Vad gäller t.ex. om du går till stranden och krossar alla fönsterrutor på vägen dit? Är denna handling tillåten (är det tillåtet att utföra denna handling)? Mer allmänt kan man ställa följande fråga. Om en typ av handling är tillåten, betyder det att det är tillåtet att utföra *varje* partikulär handling av denna typ eller betyder det att det är tillåtet att utföra *en* partikulär handling av denna typ? Eller hur skall man förstå en sådan tillåtelse?

Det är inte uppenbart hur dessa frågor bör besvaras och jag tänker inte heller försöka mig på ett generellt svar i denna uppsats. Jag nöjer mig med att konstatera att om man symboliserar argument 5 (och liknande argument) på det föreslagna sättet, så kan man förklara att det (de) är giltigt (giltiga).

10. Slutsats

Den här uppsatsen har handlat om fritt val tillåtelser. Jag har gått igenom den s.k. fritt val tillåtelser paradoxen och har tagit upp några möjliga lösningar. Jag har presenterat ett eget förslag på hur man kan förstå tillåtelser av denna typ och hur man kan lösa FVT paradoxen. Jag nämnde några potentiella invändningar mot denna analys och visade hur dessa kan bemötas.

Den generella tanken bakom mitt förslag är att fritt val tillåtelser ofta kan tolkas på följande sätt. Det finns en mängd entiteter och för varje entitet i denna mängd gäller det att du fritt får välja om du skall förhålla dig till denna entitet på ett visst sätt eller inte. T.ex.: Det finns en mängd ting och för varje ting i denna mängd gäller det att du fritt får välja om du skall ta det eller inte, ge bort det eller inte, köpa det eller inte, sälja det eller inte, låna det eller inte, låna ut det eller inte osv. Det finns en mängd platser och för varje plats i denna mängd gäller det att du fritt får välja om du skall gå till den eller inte, resa till den eller inte osv. Det finns en mängd personer och för varje person i denna mängd gäller det att du fritt får välja om du skall hjälpa denna person eller inte, flytta ihop med denna person eller inte, gifta dig med denna person eller inte osv. Det finns en mängd ämnen och för varje ämne i denna mängd gäller det att du fritt får välja om du skall studera det eller inte, skriva en uppsats om det eller inte osv. Det finns en mängd sånger och för varje sång i denna mängd gäller det att du fritt får välja om du skall framföra den eller inte. Det finns en mängd tomma parkeringsplatser och för varje tom parkeringsplats gäller det att du fritt får välja om du skall parkera din bil där eller inte. Det finns en mängd (möjliga) handlingar (av en viss typ) och för varje (möjlig) handling (av denna typ) i denna mängd gäller det att du fritt får välja om du skall utföra den eller inte. Osv.

Ibland har FVT paradoxen använts som ett argument emot s.k. standard deontisk logik (SDL). Jag har försökt visa att man kan acceptera förekomsten av FV tillåtelser utan att behöva förkasta SDL. Däremot pekar förekomsten av FV tillåtelser på behovet av en kvantifierad deontisk logik.

Låt mig avsluta med att kort nämna några fördelar med den lösning på FVT paradoxen jag har föreslagit.

(i) Vi kan visa att intuitionen att slutsatsen följer ur premissen i argument 1 (och liknande argument) är sann (om satserna tolkas på ”rätt” sätt). Vi kan också visa hur vissa tolkningar av argumentet gör det ogiltigt. Vi kan pragmatiskt förklara när det är rimligt att tolka argumentet på det ena sättet och när det är pragmatiskt rimligt att tolka det på det andra.

(ii) Vi behöver inte anta att ”eller” är mångtydigt eller har en icke-klassisk logik.

(iii) Vi behöver inte anta att ”eller” i fritt val tillåtelser tolkas idiomatiskt och egentligen betyder detsamma som ”och”.

(iv) Vi behöver inte lägga till några dolda operatorer för att förklara giltigheten hos våra argument.

(v) Vi behöver inte införa några nya primitiva eller definierade fritt val operatorer.

(vi) Vi behöver inte anta att fritt val tillåtelser endast ”opererar” på mängder av satser.

(vii) Vi behöver inte göra en distinktion mellan starka och svaga tillåtelser.

(viii) Många har argumenterat för en pragmatisk lösning på problemet med fritt val tillåtelser. Det är nog riktigt att ”Du får A eller B” pragmatiskt implicerar ”Du får A” och ”Du får B” i många situationer. Men om argumenten i den här uppsatsen är riktiga, så kan det förklaras med att sådana argument också kan vara semantiskt giltiga. Pragmatiska faktorer och olika satsers kontext kan också användas för att avgöra när det är rimligt att tolka en viss sats på ett sätt som gör ett argument som innehåller en disjunktiv tillåtelse (t.ex. argument 1) giltigt och när det inte är det.

(ix) Vi kan förklara att följande argument är giltigt. Premiss: Ingen får ta den här tårtbiten eller den här bakelsen. Slutsats: Ingen får ta den här tårtbiten och ingen får ta den här bakelsen. Låt Txy stå för ”x tar y”, t för den här tårtbiten och b för den här bakelsen. Följande formella argument är giltigt i alla system av det slag som beskrivs i Rönnedal (2014). Premiss: $\neg\exists xP(Txt \vee Txb)$. Slutsats: $\neg\exists xPTxt \wedge \neg\exists xPTxb$. Åtminstone några andra lösningar på (FVT) paradoxen har problem med att förklara giltigheten både hos argument av denna typ och argument av samma typ som argument 1.

(x) Vi behöver inte utveckla någon ny och komplicerad form av deontisk logik, t.ex. en dynamisk deontisk logik, en icke-monotonisk deontisk logik, eller en deontisk logik baserad på någon form av ”linjär logik” eller ”inkvisitiv” semantik. Det kan förstås finnas andra skäl att utveckla sådana logiker. Men om argumenten i den här uppsatsen är hållbara, så är fritt val tillåtelser paradoxen inte ett sådant skäl.

(xi) Vi kan undvika alla de problematiska konsekvenser som följde av att anta att $P(A \vee B)$ medför PA och PB i SDL.

(xii) I synnerhet gäller det att om argumenten i den här uppsatsen är riktiga, så kan förekomsten av fritt val tillåtelser inte användas för att visa att SDL är inkorrekt och måste överges. Det räcker med att vi kombinerar SDL med (en temporal aletisk-deontisk) predikatlogik.

För att kunna bevisa våra argument krävs det dock att SDL kombineras med predikatlogik. Detta talar för att vi behöver en kvantifierad deontisk logik. Jag har tidigare argumenterat för att det finns andra skäl att omfamna en sådan logik. Om argumenten i den här uppsatsen är riktiga, ger de ytterligare stöd för detta påstående.

Referenser

Aher, M. (2012). Free Choice in Deontic Inquisitive Semantics. In Aloni, Kimmelmann, Roelofsen, Sassoon, Schultz, Westera (red.), *Logic, language and meaning*, 18th Amsterdam Colloquium, Amsterdam, The Netherlands, December 19-21, 2011, ss. 22-31.

Aloni, M. (2007). Free choice, modals, and imperatives. *Natural Language Semantics*, 15, ss. 65–94.

Anglberger, A. J. J., Dong, H. & Roy, O. (2014). Open Reading without Free Choice. I Cariani, F., Grossi, D., Meheus, J, Parent, X. (red.). *Deontic Logic and Normative Systems*. DEON 14. Springer, ss. 19-32.

Asher, N. & Bonevac, D. (2005). Free Choice Permission Is Strong Permission. *Synthese*, Vol. 145, Nr. 3, ss. 303-323.

Barker, C. (2010). Free choice permission as resource-sensitive reasoning. *Semantics & Pragmatics*, vol. 3, ss. 1-38.

Chemla, E. (2009). Universal implicatures and free choice effects: experimental data. *Semantics & Pragmatics*, Vol. 2, Article 2, ss. 1–33.

Dignum, F. Meyer, J. -J. Ch. & Wieringa, R. J. (1996). Free Choice and Contextually Permitted Actions. *Studia Logica*, Vol. 57, Nr. 1, ss. 193-220.

Fox, D. (2007). Free choice disjunction and the theory of scalar implicature. I Uli Sauerland & Penka Stateva (eds.). *Presupposition and implicature in compositional semantics*. Palgrave MacMillan, ss. 71-120

Franke (2010). Free Choice from Iterated Best Response. I Aloni, M, Bastiaanse, H, de Jager, T, Schulz, K (red.), *Logic, Language and Meaning*: 17th Amsterdam Colloquium 2009, revised selected papers, ss. 295–304.

Föllesdal, D. & Risto, H. (1971). Deontic Logic: An Introduction. I R. Hilpinen (red). *Deontic Logic: Introductory and Systematic Readings*. Dordrecht and Boston: Reidel, ss. 1-35.

Gabbay, D., Gammaitoni, L., Sun, X. (2014). The paradoxes of permission an action based solution. *Journal of Applied Logic*, 12, ss. 179-191.

Gabbay, D., Horty, J., Parent, X., van der Meyden, E. & van der Torre, L. (red.). (2013). *Handbook of Deontic Logic and Normative Systems*. College Publications.

Geurts, B. (2005). Entertaining Alternatives: Disjunctions as Modals. *Natural Language Semantics*, 13, ss. 383–410.

Geurts, B. & Pouscoulous, N. (2009). Free choice for all: a response to Emmanuel Chemla. *Semantics & Pragmatics*, Vol. 2, Article 5, ss. 1–10.

Giannakidou, A. (2001). The Meaning of Free Choice. *Linguistics and Philosophy*, Vol. 24, Nr. 6, ss. 659-735.

Hansson, S. O. (2001). *The structure of Values and Norms*. Cambridge University Press.

Hansson, S. O. (2013). The Varieties of Permission. I D. Gabbay, J. Horty, X. Parent, E. van der Meyden & L. van der Torre (red.). *Handbook of Deontic Logic and Normative Systems*. College Publications, ss. 195-240.

Hilpinen, R. (red.). (1971). *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel Publishing Company.

Hilpinen, R. (red.). (1981). *New Studies in Deontic Logic Norms, Actions, and the Foundation of Ethics*. Dordrecht: D. Reidel Publishing Company.

Jennings, R. E. (1985). Can There Be a Natural Deontic Logic? *Synthese*, Vol. 65, Nr. 2, ss. 257-273.

Kamp, H. (1973). Free Choice Permission. *Proceedings of the Aristotelian Society*, New Series, Vol. 74, ss. 57-74.
Makinson, D. (1984). Stenius' approach to disjunctive permission. *Theoria*, 50, ss. 138-147.
Merin, A. (1992). Permission Sentences Stand in the Way of Boolean and Other Lattice-Theoretic Semantices. *Journal of Semantics*, 9, ss. 95-162.
Nute, D. (1985). 'Permission'. *Journal of Philosophical Logic*, vol. 14, ss. 169-190.
Parks, R. Z. (1973). Note on an Argument of Von Wright's. *Philosophical Studies*, Vol. 24, Nr. 1, s. 64.
Rönnedal, D. (2010). *An Introduction to Deontic Logic*. Charleston, SC.
Rönnedal, D. (2012). Temporal alethic-deontic logic and semantic tableaux. *Journal of Applied Logic*, 10, 2012, ss. 219-237.
Rönnedal, D. (2012b). *Extensions of Deontic Logic: An Investigation into some Multi-Modal Systems*. Department of Philosophy, Stockholm University.
Rönnedal, D. (2014). Quantified Temporal Alethic-Deontic Logic. *Logic and Logical Philosophy*. DOI: 10.12775/LLP.2014.016. Published online 2014.
Saeboe, K. J. (2001). The Semantics of Scandinavian Free Choice Items. *Linguistics and Philosophy*, Vol. 24, Nr. 6, ss. 737-787.
Schulz, K. (2005). You may read it now or later: A Case Study on the Paradox of Free Choice Permission. Masters Thesis, University of Amsterdam.
Schulz, K. (2005b). A Pragmatic Solution for the Paradox of Free Choice Permission. *Synthese*, Vol. 147, Nr. 2, ss. 343-377.
Simons, M. (2005). Dividing Things Up: The Semantics of Or and the Modal/Or Interaction. *Natural Language Semantics*, 13, ss. 271–316.
Stenius, E. (1982). Ross' paradox and well-formed codices. *Theoria*, 48, ss. 49-77.
Van Rooij, R. (2006). Free Choice Counterfactual Donkeys. *Journal of Semantics*, 23, ss. 383–402.
Van Rooy, R. (2000). Permission to Change. *Journal of Semantics*, 17, pp. 119-145.
Von Wright, G. H. (1968). *An Essay on Deontic Logic and the General Theory of Action*. Amsterdam: North-Holland Publishing Company.
Von Wright, G. H. (1971). Deontic Logic and the Theory of Conditions. I R. Hilpinen (red). *Deontic Logic: Introductory and Systematic Readings*. Dordrecht and Boston: Reidel, ss. 159-177.
Woleński, J. (1980). A Note on Free Choice Permissions. *Archiv für Rechts- und Sozialphilosophie / Archives for Philosophy of Law and Social Philosophy*, Vol. 66, H. 4, ss. 507-510.

Zimmermann, T. E. (2000). Free Choice Disjunction and Epistemic Possibility. *Natural Language Semantics*, 8, ss. 255–290.
Åqvist, L. (1965). Choice-Offering and Alternative-Presenting Disjunctive Commands. *Analysis*, Vol. 25, Nr. 5, ss. 182-184.
Åqvist, L. (1987). *Introduction to Deontic Logic and the Theory of Normative Systems*. Naples: Bibliopolis.
Åqvist (2002). Deontic Logic. In Gabbay & Guenthner (red.) *Handbook of Philosophical Logic*, 2nd Edition, vol. 8, Dordrecht/Boston/London: Kluwer Academic Publishers, ss. 147-264.

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

A Coherent and Comprehensible Interpretation of Saul Smilansky's Dualism

Sofia Jeppsson

Abstract
Saul Smilansky's theory of free will and moral responsibility consists of two parts; dualism and illusionism. Dualism is the thesis that both compatibilism and hard determinism are partly true, and has puzzled many philosophers. I argue that Smilansky's dualism can be given an unquestionably coherent and comprehensible interpretation if we reformulate it in terms of pro tanto reasons. Dualism so understood is the thesis that respect for persons gives us pro tanto reasons to blame wrongdoers, and also pro tanto reasons not to blame them. These reasons must be weighed against each other (and against relevant consequentialist reasons) in order to find out what we all things considered ought to do.

1. Introduction

Saul Smilansky's theory of free will and moral responsibility consists of two parts; dualism and illusionism (Smilansky 2000). Dualism is the thesis that both compatibilism and hard determinism are partly true, whereas illusionism is the thesis that people's false belief in libertarian free will (i.e., a kind of free will inconsistent with determinism, and possibly with plausible versions of indeterminism as well) has mostly beneficial effects and ought to be kept. Whether one agrees with illusionism or not (it has been questioned both whether people in general believe in libertarian free will, Nahmias 2011, and whether such a belief is beneficial, Pereboom 2014), it is not difficult to comprehend. Dualism, on the other hand, has puzzled many philosophers. Compatibilism is the thesis that free will and moral responsibility are compatible with determinism, whereas hard determinism is the thesis that they are not. If the world is deterministic, a complete description of the state of the world at any time in the past combined with a complete description of every law of nature implies any true proposition about present events, including human action. Hard determinists argue that under these circumstances, no one is really free or morally responsible for what she does. No one deserves to be blamed or punished for wrongful actions, and no one deserves to be praised or rewarded for exemplary ones. Compatibilists, on the other hand, argue that free will is still possible. There is a variety of ways compatibilist arguments can go, but the compatibilist might, for instance, stress the fact that determinism does not preclude the possibility of people acting free from compulsion and neurosis, point out that many agents in a deterministic universe still do what they have most reason to do according to their own values, and so on. Therefore, the compatibilist say, we can be free as well as morally responsible under determinism.

On the face of it, these theses seem completely opposed to each other, making it hard to see how there could possibly be a little bit of truth in both. Tamler Sommers writes that it is unclear what it means that punishments can be in one way just and in another way unjust (Sommers 2012 p. 115). Despite Smilansky's explicit claim to the contrary, Derk Pereboom thinks that Smilansky is not really a dualist at all, but simply a hard incompatibilist[1] like him (Smilansky 2000 p. 101; Pereboom 2001 pp. 130-131). James Lenman suggests that there is a distinction, in Smilansky's theory, between compatibilist justice and ultimate justice, compatibilist desert and ultimate desert and so on (Lenman 2002). In this paper, I will argue that regardless of whether dualism is true or not, it can be made unquestionably coherent and comprehensible if we reformulate Smilansky's thesis in terms of pro tanto reasons.

2. Smilansky's dualism

Smilansky argues, like so many other philosophers, that there is an intimate connection between respect for persons and justified praise and blame. Many philosophers argue that we may be obliged to blame someone for a wrongful act because we respect her as a person (e.g., Duff 1986 p. 70; Moore 1997 pp. 142-149 and p. 165; Dworkin 2011 pp. 224-225), but Smilansky focuses mostly on how respect for others demand that we do *not* blame them when they do *not* deserve to be blamed and praise them when they deserve to be praised (Smilansky 2000 pp. 19-20 and 127). However, respect for persons clearly requires that we do not blame others for events that were not under their control. Smilansky writes that there is an ethical basis for considering *every* control-undermining factor when asking whether someone deserves to be blamed (Smilansky 2000 p. 48). The relevance of some control-undermining factors is obvious; for instance, we ought not to blame people for what they did after having been drugged against their will. When we learn that people's actions were caused by less obvious factors beyond their control, such as the distant past together with the laws of nature, we cannot choose to ignore these factors merely because they are less salient than drugs

[1] A hard incompatibilist differs from a hard determinist merely in holding that we cannot have free will and moral responsibility under *in*determinism either. Smilansky, despite writing about hard *determinism*, often seem to have something like hard incompatibilism in mind.

or because taking them into account makes life too complicated. This is the kernel of truth in hard determinism. However, there is also an ethical basis for considering whether people were in control, in the *compatibilist* sense of the word, of their actions before we judge them. Some agents are in control of what they do in the sense that they rationally choose to do something for reasons of their own and then do it, whereas others act under the influence of psychosis or drugs, or simply behave irrationally due to being toddlers. To disregard the obvious differences between the first group of agents and the latter ones would be morally wrong (ibid p. 77).

Smilansky empathically stresses that he does not merely claim that we have *consequentialist* reasons to praise and blame people. There is a sense in which people actually *deserve* praise and blame, reward and punishment, in virtue of having acted with compatibilist freedom, and respect for persons requires that we take this into account when judging them (ibid pp. 27, 78, 88-89 and 101). A community where people are punished and rewarded based on whether they were in control of their actions or not in a compatibilist sense is in one way just and in another way unjust (ibid pp. 97-98). In a sense, people in this community get their just deserts, but in another sense, no one can ever deserve to be rewarded or punished.

My suggestion is that we can make more sense of Smilansky's dualism if we reformulate his claims in terms of the *reasons* we have to praise or blame, reward or punish.

3. Reasons and respect

Smilansky explicitly discusses a backward-looking kind of moral responsibility, or desert-entailing moral responsibility, both in compatibilist and incompatibilist terms. But what is desert-entailing moral responsibility, and how is it distinguished from its forward-looking counterpart? To delve deeply into this issue lies outside the scope of this paper. However, when analysing backward-looking blame, we ought to have something more to say than it being blame given for the reason that the agent deserves it, or blame given merely because the agent did something wrong. These statements do not really add anything to the claim that the blame is backward-looking. Rather, we should say that backward-looking blame is blame given with no eye to "training" the agent to behave better in the future, but rather given because we see her as a fellow moral *agent*. When I have done something wrong, I may think about the reasons against doing what I did and regret it. Likewise, when someone whom I consider a fellow moral agent has done something wrong, I may tell her the reasons against doing what she did and try to make her regret it. This is why many philosophers regard the readiness to blame others for their wrongdoing as a sign of respect; in a sense, blaming others for their wrongdoings amounts to treating them the way we treat ourselves, but unlike the way we treat little children, animals, severely psychotic people and so on (Moore 1997 pp. 142-149 and p. 165; Dworkin 2011 pp. 224-225).

Suppose, then, that another person has done something wrong – for instance, stolen my wallet when I was down at the pub. Since I ought to respect other persons, I think that I have at least a pro tanto reason to blame the thief. However, we may imagine that further information reveals that I do not have such a reason after all. Suppose, for instance, that I learn that she had been tricked into drinking more alcohol than she was aware of or drugged against her will, and in her confused state she mistook my wallet for her own. This fact shows that my supposed reason for blaming her was no real reason after all. Blaming her in this situation would show *no* respect for her; it would merely be cruel and heartless. Now, suppose instead that I learn, not that she was hopelessly confused due to alcohol or drugs, but that she grew up in a terrible family and terrible neighbourhood, had no non-criminal role models and so on. This is a more complicated situation. Respecting her as a person rather than writing her off as a hopeless case who cannot really know what she is doing does seem to give me an actual pro tanto reason to blame her. However, respecting her as a person might also require that I try to put myself in her shoes, and realize that abstaining from criminal activities might present a real challenge to her, whereas it is easy for me. The difference between us does not solely consist in me making better choices than she does; I have also been more fortunate. This might give me reason to somewhat mitigate my judgement, although I should still consider her blameworthy.

Finally, suppose that there are no circumstances that we would normally consider excusing or mitigating. Respecting the wallet thief as a person gives me a pro tanto reason to blame her. However, I realize that there is a sense in which nothing is ultimately up to us, since libertarian free will is impossible. Ultimately, everything we do, including her theft of my wallet, is caused by factors beyond anyone's control. Respecting her as a person gives me reason to take seriously the fact that the difference between us is ultimately a difference in luck; I was fortunate enough to be born with genes and exposed to environmental influences that caused me to abstain from crime, and she was unfortunate enough to be born with genes and exposed to an environment that caused her to steal my wallet.[2] As soon as we take up this "ultimate" perspective, it might be tempting to say that I was mistaken when

[2] Once again, remember that determinism is the thesis that a complete description of the state of the world at any time in the past combined with a complete description of every law of nature implies any true proposition about present events, including human action. The observation that, say, identical twins who grew up together do not behave exactly alike does not prove that the world is not deterministic, since no two people have been exposed to precisely the same environmental influences down to the last detail.

I initially thought that I had a pro tanto reason to blame her, but this is not the conclusion that Smilansky draws. Whereas it is not respectful at all, merely cruel, to blame someone for actions that she performed while hopelessly confused from alcohol or drugs that she did not even chose to consume, the same cannot be said about blaming someone for actions that she decided to perform for reasons of her own. It remains true that there *is* something disrespectful about dismissing her theft as a mere result of forces beyond her control; dismiss what she did like we might dismiss the actions of someone in the grip of psychosis, or a tiny child. Respecting her as a person therefore gives me a genuine pro tanto reason to blame her – and simultaneously another, likewise genuine, pro tanto reason not to blame her.

It might be objected that whereas it is clearly possible that I find myself in a situation where, say, respect for persons gives me a pro tanto reason to blame someone whereas the fact that blaming her would have disastrous consequences gives me a pro tanto reason not to blame her, it cannot be the case that respect for persons alone gives me both a reason to blame and a reason not to blame. But situations where the same value or general duty gives rise to two opposing pro tanto reasons are actually commonplace. Suppose that I am a parent, and have a general duty to make my child happy. This gives me a pro tanto reason to buy her candy from time to time if she loves candy, and also a pro tanto reason not to give her candy since candy is bad for her health, and good health is important for her happiness. I will end up weighing these reasons against each other, and may sometimes conclude that I all things considered ought to buy her a little bit of candy, at other times conclude that I all things considered ought to refuse her requests. In a similar vein, respect for persons may simultaneously give me a reason to take other people and their agency seriously by praising and blaming them for actions that are under their control in a compatibilist sense, and a reason to consider the fact that differences in behaviour are ultimately just due to luck and therefore not praise or blame them for anything they do.

4. How we ought to treat people all things considered

How are we to weigh these different reasons against each other? Overall, Smilansky's position seems to be that we usually have all-things-considered reasons to praise and blame people, reward and punish, despite the fact that doing so is always ultimately unjust. He further argues that the extent to which a given punishment is unjust depends in part on how harsh it is, in part on the level of compatibilist control with which the agent acted (ibid p. 134). He argues that when we have determined that there is injustice, it is a separate question whether something ought to be done about the injustice (ibid p. 135). Although he focuses on a kind of compatibilism concerned with desert and backward-looking reasons for praise and blame, there can also be important consequentialist reasons for praising and blaming, rewarding and punishing (ibid p. 102). Hard determinist considerations, however, should play a mitigating role (ibid p. 102).

Bringing all this together to a consistent theory about the reasons we have to praise, blame, punish and reward, gives us the following picture: We have respect-based pro tanto reasons to praise and blame, punish and reward people for actions they perform with compatibilist control. This is the kernel of truth in compatibilism. We also have consequentialist pro tanto reasons to create practices and institutions that administer praise, blame, punishments and rewards for this category of actions. Finally, we have respect-based pro tanto reasons never to praise, blame, punish or reward anyone. This is the kernel of truth in hard determinism. Reasons of the first kind can be affected by *how much* compatibilist control the agent had. Reasons of the second kind can be stronger or weaker depending on exactly how good or bad the consequences will be. Reasons of the third kind are, in the case of blame and punishment at least, stronger the harsher the blame or punishment is. If we punish someone harshly enough, even if she acted with compatibilist control, there will presumably be a point where the reason not to punish comes to outweigh the reasons we do have to punish, and thus we must diminish the amount of punishment we plan to inflict on her until the balance of reasons is such that punishing her once again becomes justified. In this way, hard determinist considerations serve a mitigating function. As long as we stick to social sanctions and more humane punishments, the reasons we do have to blame and punish will, insofar as wrongdoers had compatibilist control over their actions, outweigh the reasons we have not to do so.

5. Conclusion

The claim that there is truth in both (backward-looking) compatibilism and hard determinism can be understood as us having respect-based, backward-looking pro tanto reasons both *to* praise and blame, reward and punish agents for actions performed with compatibilist control, and respect-based, backward-looking pro tanto reasons *not to* do this. It is commonplace and non-mysterious that we might have a pro tanto reason to A and a pro tanto reason not to A derived from the same value or general duty, and that we must weigh these reasons against each other to find out what we ought to do all things considered. Thus, if we reformulate Smilansky's dualism in terms of reasons for praising, blaming, rewarding and punishing, we see that it is a coherent and comprehensible thesis. Whether it is also true is a topic for other papers.

Acknowledgements
I want to thank Saul Smilansky for encouraging this interpretation of his thesis in conversation.

References

Duff, A. (1986). *Trials and punishment.* Cambridge: Cambridge University Press.

Dworkin, R. (2011). *Justice for Hedgehogs.* London: The Belknap Press of Harvard University.

Lenman, J. (2002). On the Alleged Shallowness of Compatibilism: A Critical Study of Smilansky, S. 2000. *Free Will and Illusion.* Oxford: Clarendon Press.

Moore, M. (1997). *Placing Blame. A Theory of the Criminal Law.* Oxford: Oxford University Press.

Nahmias, E. (2011). Intuitions about Free Will, Determinism and Bypassing. In R. Kane, eds, *The Oxford Handbook of Free Will, 2:nd edition,* pp. 555-575. New York: Oxford University Press.

Pereboom, D. (2001). *Living Without Free Will.* Cambridge: Cambridge University Press.

Pereboom, D. (2014). *Free will, Agency and Meaning in Life.* New York: Oxford University Press.

Smilansky, S. (2000). *Free Will and Illusion*. Oxford: Clarendon Press.

Sommers, T. (2012). *Relative Justice: Cultural Diversity, Free Will and Moral Responsibility.* Princeton: Princeton University Press.

Sofia Jeppsson
Department of Philosophy
Stockholm University
sofia.jeppsson@philosophy.su.se

The Compensation Principle

William Simkulet

Abstract
In "Should Race Matter?," David Boonin proposes the compensation principle: When an agent wrongfully harms another person, she incurs a moral obligation to compensate that person for the harms she has caused. Boonin then argues that the United States government has wrongfully harmed black Americans by adopting pro-slavery laws and other discriminatory laws and practices following the end of slavery, and therefore the United States government has an obligation to pay reparations for slavery and discriminatory laws and practices to those who have been harmed by them - in particular, to contemporary black Americans. Here I argue that the compensation principle is false because it violates the control principle, the foundational principle of ethics that states that moral responsibility requires control; for an agent to be morally responsible for something, whether or not she does that thing must be within her control. If the compensation principle creates a moral obligation for an agent to compensate a harmed party, failure to do so will result in that agent's being morally blameworthy for failing in her obligation. Because some harms cannot be compensated for, agents who wrongfully harm others will be required to do something that is outside of their control.

Introduction
In "Should Race Matter?," David Boonin proposes a moral principle, the *compensation principle*:

> [I]f someone wrongfully harms another person, then he incurs a moral obligation to compensate his victim for the harms that he has wrongfully caused. (43)

On the back of this principle, Boonin constructs the compensation argument - an argument he believes shows that the United States government has a moral obligation to pay slavery reparations to contemporary black Americans. This argument fails for two reasons: First, I argue Boonin's account of the compensation principle is inconsistent with a host of commonsense, foundational principles concerning the nature of moral responsibility. Second, the compensation argument rests upon two problematic assumptions:

(i) that people are morally entitled to their relative's wealth, and (ii) that governments are morally culpable for the foreseeable harms of the laws they create.

This paper is divided into two sections. The first section is an ethical analysis of the compensation principle, the second section is a critique of the compensation argument.

1. On the Compensation Principle

To have a *moral obligation* to do a thing, *x,* is to be the appropriate object of moral blame if one doesn't do *x*, and the appropriate object of moral praise if one does *x*. Although we sometimes talk about agents succeeding or failing in regards to their moral obligations, this talk should not be understood to imply that there is an element of chance in whether one is morally blameworthy or praiseworthy. It wouldn't make sense to hold people morally accountable for matters of chance or luck. If one has a moral obligation to do a thing, *x*, then whether or not *x* occurs must be up to them. This concept is illustrated in the following principle:

> *The Control Principle* - To be morally responsible for a thing, *x*, is to have control over whether *x* occurs. (CP from now on).

While CP concerns moral responsibility, a parallel principle - what David Boonin characterizes as the the principle that "ought implies can" (81) - concerns moral obligation:

> *The "Ought Implies Can" Principle* - If an agent *a* has a moral obligation to do *x*, it is within *a*'s power to do *x*. (OICP from now on).

Because having a moral obligation to do *x* means you are blameworthy if you fail to do *x*, OICP follows from CP. To have the moral obligation to do *x* and fail is a moral failing, which is to say that it was up to you whether you would *x* or ~*x*, you freely chose the latter, and that you are blameworthy for that choice. Moral responsibility requires control in the form of legitimate alternate possibilities - if one couldn't actually do otherwise, one can't be responsible for her actions. This leads to the following principle:

> *The Principle of Alternate Possibilities* - If an agent *a* is morally blameworthy for some act *y* then *a* could have done otherwise. (PAP from now on).

Because moral responsibility requires control, it is impossible to find yourself in a situation where all of your possible options are blameworthy. Michael Otsuka (1998) captures a key feature of PAP in the following principle:

> *The Principle of Avoidable Blame* - If an agent *a* is morally blameworthy for some act *y* then there is some alternative act *z* that *a* had the power to do, and had *a* acted to bring about *z* rather than *y*, *a* would have been entirely morally blameless. (PAB from now on.)

According to PAB, to be blameworthy requires the opportunity to act in a blameless manner. But consider the following case:

> *Ethics final*- Having completed all of her coursework except for one final class - Philosophy 241 Ethics, Susan enrolls in this course and promises her parents that she will graduate. Susan freely decides to disregard her coursework and focus her attention on enjoying her final semester playing video games all day. An hour before her Ethics final exam, Susan decides to check her email for the first time in the semester and finds that she will fail the course (and fail to graduate) unless she does well on the final exam. Unprepared to take the exam, Susan believes she has only two options - cheat, or not cheat and fail the exam, fail to graduate, and break her promise.

Both options are *prima facie* morally bad and we might be tempted to say that she's blameworthy if she goes with either option. The important feature about this case, though, is that her blameworthiness can be traced back to her prior free choices to focus her attention on enjoying her final semester, instead of completing the coursework. Suppose that Susan chooses not to cheat. We might say that Susan is morally blameworthy for failing the test and breaking her promise, but any responsibility for this would be derivative moral responsibility, where an agent is *derivatively morally responsible* for *x* if and only if she is truly morally responsible for *y* and *y* was appropriately connected to *x*. Susan's moral failing here is her decision to disregard her coursework, and she could have behaved in a manner in which she would be entirely blameless in this case - she could have freely chosen not to disregard her coursework!

Consider David Boonin's compensation principle:

> *The Compensation Principle* - When an agent *a* wrongfully harms another agent *b*, *a* incurs a moral obligation to compensate *b* for those harms.

Boonin says "If I wrongfully damage your car by vandalizing it, for example, then I incur a moral obligation to repair the damage that I caused or pay the costs involved in having someone else repair it on my behalf." (43) At first glance, this principle has some intuitive force - if one harms another, it seems only fair that she should try to undo that harm. It even bears some similarity to the "you break it, you bought it" policy you might expect your local stores to have... although that policy is more strict, as you are responsible for anything you break, even if your breaking it was morally acceptable.

The problem with this principle is that it is inconsistent with CP. Consider a case where compensation is impossible:

> *The priceless vase* - Jack and Jill visit the priceless, one-of-a-kind vase display at their local museum and notice that one vase in particular is poorly mounted behind the glass. Jack contends that if they were to jump up and down and make enough vibrations, the vase would fall from the mount and be shattered. Jill disagrees, contending that there's no way jumping up and down would cause such a result, citing several books she had read on the subject. Jack maliciously begins jumping up and down, hoping to break the vase, while Jill begins jumping up and down to prove her point, believing that doing so will not break the vase. The vase breaks.

According to the "you break it, you bought it" policy, both Jack and Jill seem to be equally on the hook for breaking the vase... but the vase is not for sale, and there is no way they can replace the priceless, one-of-a-kind vase. If Jack and possibly Jill acquired the moral obligation to replace the vase, CP, OICP, and PAB are false. Jack and Jill would be morally obligated to replace the vase, but replacing the vase would be outside of their control; it's something that they just can't do. Although both Jack and Jill might have a list of alternate possibilities (lying, running away, apologizing, suicide, etc.) because none of these possible futures will result in adequate compensation, Jack and possibly Jill will fail in any moral obligation to compensate the museum, and will thus not be able to avoid blame for this moral obligation.

There is another problem with the compensation principle. One of the virtues of the compensation principle is that it explains our intuition that Jack has a moral obligation to try to clean up his mess because he intentionally created the mess; however what about Jill? It strikes me that Jill should feel bad about what happened, and that if it was possible to replace the vase, she would want to do so. If the compensation principle is supposed to explain why Jack incurs a moral obligation to replace the vase, shouldn't it also be able to explain why Jill might believe she has an obligation to replace the vase?

Suppose, for a moment, that Jill didn't have this intuition after breaking the vase. Suppose a security guard approaches just as the vase breaks, and Jack laughs and runs away hoping to avoid getting caught. Imagine the guard asks Jill what she has to say for herself, and that Jill replied "I'm sorry for the museum's loss. Of course had I known that jumping up and down could lead to the vase breaking, I would never have done it. But I didn't know this, and my actions were completely innocent - I freely engaged in an action that I was justified in believing would have no bad consequences, and therefore it is not my fault that the vase broke and I have no reason to compensate you for your loss, despite the fact that I caused it." I imagine you'd find this response unsatisfactory.

If it makes sense to say that Jill would have some feeling of responsibility with regards to the vase, then the compensation principle is unsatisfactory because it seeks to explain Jack's obligations with regards to the vase, but not Jill's. Suppose that the intuitive force behind the compensation principle and the "you break it, you buy it," policy has nothing to do with the agent and their intentions, but rather with the actual harm produced:

> *The plain vase* - Jack and Jill promise their neighbors that they will water their plants while they are gone. One day, while doing so, they notice a plain, ordinary vase - the kind one could find at any flower shop for a low price - perched precariously on the edge of the kitchen counter. Jack contends that if they were to jump up and down and make enough vibrations, the vase would fall from the counter and be shattered. Jill disagrees, contending that there's no way jumping up and down would cause such a result, citing several books she had read on the subject. Jack maliciously begins jumping up and down, hoping to break the vase, while Jill begins jumping up and down believing that doing so will not break the vase. The vase breaks.

Suppose that after breaking the vase, Jill has the intuition that she should replace the vase, and she does so because of this. Her goal here is to make up for the harm that she has inadvertently caused to her neighbors, not to pay for some moral debt that she has incurred.

Jack willfully and maliciously sought to break the vase, so it makes sense to say that he's blameworthy for this action. Boonin's compensation principle contends that because his wrongful actions harmed his neighbors, he incurs a moral obligation to compensate them. The problem is that Jill also believes that she has an obligation to compensate them, despite her actions not being malicious. This suggests that the wrongful nature of Jack's action has nothing to do with our intuition that he ought to replace the vase.

Suppose, though, that rather than a traditional moral obligation, we're to understand the incurred debt of Boonin's compensation principle is a means to free one from one's blameworthiness. In both of the vase cases, Jack has acted in a *prima facie* immoral way - he willfully and maliciously sought to destroy the property of someone else. In both cases, Jack could have done otherwise - he could have refrained from activity designed to destroy the respective vases. In both cases, this was within his control, and in both cases, had he refrained, he would have been morally blameless. When Jack acts to destroy the vases, in both cases he is morally blameworthy. If Jack were to replace the vase in the latter case, he would negate the harm caused by his wrongful act. Maybe the intuition behind the compensation principle is that by providing compensation we are absolving ourselves of our previous wrongdoing. Although Jill has harmed others in these cases, she has not done so intentionally. Therefore she is not morally culpable for breaking the vases, and therefore even if she did compensate the museum or her neighbors for her actions, this would not absolve her of any moral wrongdoing because she hasn't done anything wrong!

If the moral obligation at play in the compensation principle is of this nature, then it isn't a moral obligation in the traditional sense because failure to live up to the obligation doesn't create an additional moral stain on the agent. This account of the compensation principle would, then, be compatible with CP and OICP. We might call this a derivative moral obligation, where one has a derivative moral obligation to perform some action x if and only if x will erase their responsibility for some previous action y. The concept of a derivative moral obligation is appealing because it will allow wrongdoers to actually make up for their previous wrong actions. However, consider the following case:

> *Destroying Pie* - Frank wants to do something nice for his friend Carol, and spends all afternoon baking her a pie for desert. Frank gives Carol the pie, then inexplicably Frank takes the pie back and throws it into the garbage, sets the garbage on fire, and throws the garbage bag into the trash - utterly destroying the pie.

Because of the apparent symmetry between praise and blame, this might be what we could call a derivative moral freedom, where one has a *derivative moral freedom* to perform some action x if and only if x will erase their praise for some previous action y. It makes sense to say that Frank's actions actually undo the praiseworthiness for his previous actions.

The primary problem with the concepts of derivative moral freedom and derivative moral obligation is that they ask us to consider distinct moral actions in unison. Frank's pie baking is *prima facie* morally praiseworthy, but his pie destruction erases the good created by his pie baking. We're expected to evaluate the actions together, rather than separately, but this easily leads to an unintuitive picture of morality. If the concept of derivative moral obligation makes sense, then immoral actions can be done with impunity and later removed from our moral records to restore virtue. Suppose that every day at noon Jack freely and maliciously breaks his neighbor's lawn gnome, and every day at 2pm he replaces the lawn gnome with a copy from the local garden shop. On this account of the compensation principle, Jack is blameworthy from 12pm-2pm, but blameless from 2pm-12pm. This is absurd! Jack acts in a villainous manner without regard for the property of others and this is *prima facie* morally blameworthy even with compensation! If the compensation principle is true, then his neighbors, upon discovering his routine, should not judge him morally blameworthy after 2pm! But this would throw property rights out the window. Anyone would have full and unfettered access to anyone else's property, morally, so long as it doesn't result in any permanent harm!

Another possibility is that derivative moral obligations don't completely wipe away moral blame. If Boonin were to vandalize his neighbor's car, for example, he may incur the derivative moral obligation to pay for someone else to fix it, but he might also retain some aspect of his original blameworthiness - for examples, his neighbors might be justified in still blaming him, but they shouldn't blame him as much as if he were had vandalized the car and not had it fixed. However, if this is case, I'm not so sure we want to say that one incurs a moral obligation, derivative or otherwise, to repair the harm one has caused from one's misdeeds, rather we might say that an appropriate punishment for one's previous action is, in part, to repair the harm. I suppose one could talk about a moral obligation to accept one's moral punishment gracefully, but that is a radically different principle than Boonin's compensation principle seems to be.

It is also unclear how this solution will adequately explain the intuition that Jill has that she should try to compensate for the harm she caused. The benefit of derivative moral obligation talk is that it seems to allow us to get around CP, OICP, and PAB. It's a matter of luck whether Jack will be able to compensate others, but compensating isn't praiseworthy; rather it's blame-erasing. Because Jill isn't blameworthy for her action, compensation can't erase any blame. Still, if Jill believes she ought to compensate her neighbors for breaking their vase, it seems she would be best described as believing she has a moral obligation to do so, not a derivative moral obligation. This leads to several odd disparallels between Jack and Jill, most notably that Jill would incur a moral obligation that Jack doesn't by acting in a morally responsible manner (by unexpectedly and unintentionally breaking the vase, where Jack intentionally and maliciously broke the vase).

The compensation principle doesn't seem to pick up on a steadfast moral principle, rather it strikes me that it's best understood as a rule of thumb to govern appropriate responses to either advertent or inadvertent harm.

> *The Compensation Principle** - When an agent *a* causes a harm to another agent *b*, *a* incurs an *ad hoc* moral obligation to compensate *b* for said harm if and only if (i) it is possible to compensate *b*, (ii) *b* did not act in such a way as to deliberately allow *a* to harm *b*, and (iii) compensating *b* does not constitute an overwhelming burden to *a*.

Rather than dealing with moral obligations at large, this principle is concerned with *ad hoc* moral obligations, understood here as moral obligations dealing with compensating harm to others. Immoral actions might also produce additional moral obligations unrelated to compensating either actual or possible harm. Here I've introduced qualification (i) to ensure that this principle is compatible with CP. Furthermore, because a moral principle should not encourage or incentivize immoral actions, (ii) prevents unscrupulous shopkeepers from placing their vases precariously on the edge of tables under the sign "You break it, you bought it." Finally, (iii) strikes me as a reasonable limitation to account for cases like this:

> *The expensive vase* - Jack volunteers to water his neighbor's plants, and after he does so, he notices his neighbor's have a nice looking vase. Upset at the volume of plants he has to water, he willfully and maliciously takes his frustrations out on the vase, throwing against the wall, shattering it. He immediately regrets his actions, giving into his frustration, and offers to replace it when his neighbors come home. His neighbors tell him the vase costs more than Jack makes in a decade.

My intuition is that in this case Jack is morally blameworthy for his action, but despite his moral intuition that he should compensate his neighbors, he lacks a moral obligation to do so because to compensate them for the broken vase would constitute a significant burden to Jack vastly disproportionate to his immoral action. Even if Jack acted to destroy the vase knowing it was expensive, and to purposely deprive his neighbors of the vase, my intuition is he would not have even an *ad hoc* moral obligation to replace the vase. In such cases, full compensation would be too much of a moral burden. However, it strikes me that in some cases, full compensation might be insufficient. When Jack breaks his neighbor's cheap lawn gnome, he might incur the obligation to replace it, but he might be obligated to do more as well.

What makes an action morally wrong is its epistemic nature - actions done intentionally to cause harm, or without regard to the welfare of others are *prima facie* wrong, and their wrongness can be directly traced back to the beliefs and desires the agent acted on and with. It is unsurprising that malicious or negligent actions can have actual, harmful consequences. However, as illustrated by the difference in harm between the priceless vase case and the plain vase case, the consequences of one's actions are not necessarily, or even often, reflective of the immorality of the action in question - in both cases, Jack acts immorally, and Jill morally, and yet both cause harm, and far different harm in the former case than the latter. To tie moral responsibility, or even *ad hoc* moral obligations such as the ones the compensation principle deals with, to the actual value of a contingently replaceable item would be both arbitrary and, in cases like the expensive vase case, inappropriately harsh. Keep in mind the interpretation of the principle I sketch above would create moral obligations for agents, like Jill, who are entirely morally blameless for the harm their actions cause.

The compensation principle is an odd moral principle, reflecting our moral desire to not harm others, rather than offering a substantive account of what makes an agent blameworthy or praiseworthy for their actions. Compensating victims is not a get-out-of-jail-free card, as Boonin's interpretation might lead one to believe; rather the obligation to compensate others for the harms one causes seems to come more out of fear and disgust at the harms themselves, rather than being a form of moral penance. Jill offers to replace the plain vase not because she broke it, but because she is sorry that it was broken. Jack ought to offer to replace the plain vase to lessen the actual harm created by his independently blameworthy action. To fail to do so is to fail to appreciate the harm caused - and this would be an additional, separate moral failing on Jack's part, independent from his blameworthy actions.

2. On the Compensation Argument

For Boonin, the compensation principle serves as the backbone for what he calls the compensation argument:

The Compensation Argument

1. *Boonin's Compensation Principle* - If a party wrongfully harms someone, that party incurs a moral obligation to compensate that someone.
2. The United States government wrongfully harmed past generations of black people by allowing slavery and discrimination.
3. These harms continue to harm black Americans today.
4. The government's obligation has not been met, and persists. (53-54)
   Conclusion:

Therefore, the United States government has a moral obligation to provide compensation to currently living black Americans.

Before evaluating this argument, it'll be useful to evaluate two, simpler arguments that Boonin might have advocated immediately after the abolition of slavery.

The Individual Compensation Argument

1. *Boonin's Compensation Principle*
2. Slavers, slave owners, and many individuals wrongfully harmed current and past generations of black people by enslaving them and treating them as property rather than with the respect they deserve.

   Conclusion:

   Therefore, these individuals have a moral obligation to compensate *at least*[1] current generations of former-slaves for the harms committed.

The Government Compensation Argument

1. *Boonin's Compensation Principle*
2. The United States government wrongfully harmed current and past generations of black people by slavery and discrimination.

   Conclusion:

   Therefore, the government has a moral obligation to compensate *at least* current generations of former-slaves for the harms committed.[2]

Which of these arguments is more persuasive? It strikes me that the former is more persuasive than the latter - Slave owners uncontroversially owed a substantial moral debt to their former slaves, and this is a debt so massive that it cannot be fully paid. There is no adequate compensation for slavery; retroactively paying slaves even the most generous of wages would fail to adequately compensate them for the loss of their autonomy.

The unforgivable and uncompensatable nature of slavery serves to remind us that Boonin's compensation principle is false because of its inconsistency with CP and OICP. Slave owners can't adequately compensate slaves for the harms inflicted on them, therefore they can't have the obligation to do so! There is another principle that might offer a better explanation of the slave owners' obligations:

---

[1] Proponents of this argument would advocate for some form of compensation for previous generations as well; however with regards to evaluating this argument I will set aside this issue because it's not clear how such compensation would work.

[2] Note that the government did attempt to provide some former slaves with some compensation, but this was clearly woefully inadequate compensation given the grievous injustices of slavery and discrimination.

The Theft Principle - When *a* steals *x* from *b*, *x* belongs to *b*, not *a*. *x*, and/or all proceeds derived from *x* belongs to *b*, not *a*.

The rational for the theft principle is clear - it is uncontroversially true that stolen property does not become the property of the thief, regardless of how it looks to ignorant third party observers. Theft can be understood as unjust restriction of access to the personal property of others. Second, it is uncontroversially morally true that moral principles should never incentivize immoral activity. If proceeds from ill gotten gains belonged to the thief, the thief would have practical incentives to steal and invest wisely. When caught such a thief could return the stolen goods, offer herself up for punishment, and then walk away with the proceeds.[3] The theft principle builds in the idea that proceeds, too, belong to the victim - there is no more appropriate party to benefit than the wronged party.

Slavery includes the theft of the slave's labor and property. This leads to the following arugment:

The Theft Argument

1. *The Theft Principle*
2. Slavers and slave owners stole the labor and property of black persons.

   Conclusion:

   Therefore, these individuals do not own the goods they stole, nor do they own the wealth built on the back of these goods. Their former slaves own this wealth.

The theft argument strikes me as a more plausible argument than the individual compensation argument, if only because it is consistent with CP and OICP. Nothing can properly compensate former slaves for the harms that have been inflicted on them, but it is at least fair to recognize that their tormenters do not own the property they claim to own, and to return fair access to this property to the former-slaves it was stolen from, or built from.

It would have been relatively easy to follow through with the conclusion of the theft argument when slavery was abolished. Unfortunately, however, property rightfully owned by many former slaves was not returned. Applying the theft principle today is more challenging since neither slave owners or former slaves are still alive.

Boonin's compensation argument is designed to show how the debt in question has not been swept away by the ravages of time. First, he contends, like in the government compensation argument, that the government has an obligation to those harmed by their unjust social practices. Second, he

[3] A similar argument can be made for the moral support of abortion in rape cases.

contends that the present day economic inequality of black Americans compared to white Americans is the result of slavery, past discrimination, and the failure to appropriately compensate past generations of black Americans.

Boonin contends that the relative lack of welfare for black Americans today is evidence that they have been harmed by the government. (17) The analysis seems to go something like this - if past black persons had been adequately compensated for slavery, they would have likely had substantially more wealth to pass on to future generations. Thus, the United States government has harmed present day black Americans by failing to compensate past generations whose wealth they would inherit, and thus depriving them of the wellbeing they would likely have had if the government had acted morally. What is odd about this position is that it *seems* to say that children are morally entitled to the wealth of their parents, but surely this is not correct.

More troubling, though, is the rationale Boonin has for the origin of the government's debt. He contends that a government has wrongfully harmed a party if the harm is a foreseeable consequence of the government's social policies - in this case, the government allowed slavery, and allowed discrimination, and thus is morally culpable for that. According to the compensation principle, the government incurs a substantial moral debt - the debt to repay *all* injustices that were legally committed by its citizens. Set aside, for a moment, the tremendous moral debt from slavery, Native American relocation, the internment of Japanese Americans during World War II, and various laws and practices that currently, or in the past, allowed for discrimination against women or any other group based off of arbitrary social practices. Consider the harm created by the government's protection of free speech. It is uncontroversially true that allowing free speech causes substantive moral harms - for example, sexist and racist attitudes are often affirmed, and these views can contribute to significant social inequality as well as provoke moral outrage. However, it is equally uncontroversially true that free speech plays an important social regulatory role, and that to infringe upon it is to cause far worse injustices. Boonin's application of the compensation principle to the government would hold it accountable for the harms created as a result of free speech, despite the moral righteousness of allowing free speech. As such, Boonin's compensation principle would be at odds with the PAB - he would hold that the government can act in a morally responsible manner, and still be morally blameworthy for the negative consequences of its justified action.

One might object that the government doesn't wrongfully harm people through free speech, individual speakers do. The problem is that the same can be said for slavery laws - the government doesn't wrongfully harm individuals by allowing slavery, the slavers wrongfully harm those individuals. What makes the pro-slavery laws unjust is not that people are harmed by their existence; it's that they are laws that allow other people to act in blatantly immoral ways, usurping the liberty and property of others without fear of prosecution or penalty. Such laws would be unjust even if no one in the United States had ever owned slaves.

Finally, even if you were persuaded by Boonin's compensation argument, I've argued that the compensation principle is inconsistent with our most foundational moral principles - CP and OICP. Even the revised compensation principle I advocated in the previous section would fail to show that the government owes the substantial economic debt Boonin's argument would imply the government owes to black, Japanese, Native, women, and disabled Americans, among others.

There is hope, though, for reparations-advocates. The conclusion of the theft argument is that slavers and slave owners did not own their ill-gotten gains. If this is true, they could not morally bequeath them onto their heirs. This suggests that a good portion of the wealth owned by descendants of slave owners rightfully belongs to past generations of slaves. Because the past generations of black Americans would probably have passed it on to their families, it seems fair to recognize that ownership of this wealth rightfully belongs to their relatives. Note, though, that because one is not blameworthy for the actions of their ancestors, it may be inappropriate to give the proceeds subsequent generations have made on the back of this ill-gotten wealth to the ancestors of the slaves in question. Instead, it seems fair that the wealthy heirs of former slave-owning families should pay a modest, *ad hoc* interest on the money they inadvertently borrowed. To recognize the appropriate ownership of wealth in such cases will likely cause strife to the wealthy descendants of slave owners, but the injustice to them was done on the part of their ancestors who wrongfully deeded them what did not belong to them.

It is worth noting that the wealth that should be returned to the families of former slaves is far from being adequate compensation for slavery - there can be no compensation for such a horrible crime, and the returning of this wealth would do little to rectify the social inequalities that have resulted from slavery and discrimination. However, this should not come as a surprise, as immoral actions tend to create harm. Acting in accordance with the theft principle doesn't negate all of the harms caused by the injustices black Americans have faced - it only returns whatever it can to rightful owners.

Conclusion

Here I've argued that Boonin's compensation principle is inconsistent with the control principle, creating a moral obligation that in many cases can never be paid. According to Boonin's compensation argument, the evidence that the

United States has not properly compensated black Americans for slavery is their relative lack of welfare compared to white Americans - but even elevating black Americans to well beyond the economic status of white Americans would fail to adequately compensate these people for the injustices they and previous generations have faced. Such injustices taint America's history, and reparations of the kind Boonin advocates are insufficient, inappropriate, and misplaced. However, the theft principle offers a commonsense approach to dealing with one particular injustice that is a result of slavery - the theft of the labor and other property by slave owners. By returning these goods, and the financial successes built upon these goods, to the descendants of former slaves, one acknowledges the property rights of black Americans in their person and labor. The moral blight of slavery can never be removed from the history of the United States, but the continued denial that former slaves owned the fruits of their labor can, and should, be ceased.

References

Boonin, David (2011). *Should Race Matter? Unusual Answers to the Usual Questions*. Cambridge: Cambridge University Press.

Otsuka, Michael (1998). Incompatibilism and the Avoidability of Blame. *Ethics* 108(4), pp. 685-701.

William Simkulet
University of Wisconsin - Marshfield/Wood County
Simkuletwm@yahoo.com

Alethic-Deontic Logic: Some Theorems

Daniel Rönnedal

Abstract
The purpose of this paper is to prove some theorems in alethic-deontic logic. Alethic-deontic logic is a kind of bimodal logic that combines ordinary alethic (modal) logic and deontic logic. Ordinary alethic logic is a branch of logic that deals with modal concepts, such as necessity and possibility, modal sentences, arguments and systems. Deontic logic is the logic of norms. It is about normative words, such as “ought”, “right” and “wrong”, normative sentences, arguments and systems. Alethic-deontic logic contains both modal and normative concepts and can be used to study how these interact. This paper contains some interesting theorems that can be proved in alethic-deontic logic. I will show that all primitive deontic operators are redundant when prefixed to the alethic operators in some systems. I will prove that necessarily equivalent sentences have the same deontic status in many systems. I will establish that the set of sentences in some alethic-deontic systems can be partitioned into five, mutually exclusive, exhaustive subsets. Finally, I will show that there are exactly ten distinct modalities in some alethic-deontic systems.

1. Introduction

The purpose of this paper is to prove some theorems in alethic-deontic logic. Alethic-deontic logic is a kind of bimodal logic that combines ordinary alethic (modal) logic and deontic logic. Ordinary alethic logic is a branch of logic that deals with modal concepts, such as necessity and possibility, modal sentences, arguments and systems. For some introductions, see e.g. Chellas (1980), Blackburn, de Rijke, & Venema (2001), Blackburn, van Benthem, Wolter (eds.) (2007), Fitting & Mendelsohn (1998), Gabbay (1976), Gabbay & Guenthner (2001), Kracht (1999), Garson (2006), Girle (2000), Lewis & Langford (1932), Popkorn (1994), Segerberg (1971), and Zeman (1973). Deontic logic is the logic of norms. It deals with normative words, such as “ought”, “right” and “wrong”, normative sentences, arguments and systems. Introductions to this branch can be found in e.g. Gabbay, Horty, Parent, van der Meyden & van der Torre (eds.) (2013), Hilpinen (1971), (1981), Rönnedal (2010), and Åqvist (1987), (2002). Alethic-deontic logic contains both modal and normative concepts and can be used to study how these interact. In the paper Rönnedal (2012) I say more about various bimodal

Filosofiska Notiser, Årgång 2, Nr. 1, Februari 2015, 61-77

systems. Alan R. Anderson was perhaps the first philosopher to combine alethic and deontic logic (see Anderson (1956)). Fine & Schurz (1996), Gabbay & Guenthner (2001), Gabbay, Kurucz, Wolter, Zakharyaschev (2003), Kracht (1999), and Kracht & Wolter (1991) include more information about how to combine various logical systems.

This paper contains some interesting theorems that can be proved in alethic-deontic logic. It is divided into 6 sections. Section 2 contains an introduction to the systems we study in this paper. In section 3 I will prove that all primitive deontic operators are redundant when prefixed to the alethic operators in some systems. Section 4 includes a proof of the fact that necessarily equivalent sentences have the same deontic status in many systems. In section 5 I will establish that the set of sentences in some alethic-deontic systems can be partitioned into five, mutually exclusive, exhaustive subsets. Finally, in section 6, I will prove that there are exactly ten distinct modalities in some alethic-deontic systems.

2. Alethic-deontic logic

In this section I will briefly describe the alethic-deontic logics that we will study in this essay. In the paper Rönnedal (2012) I say more about them and about bimodal systems in general. For more background information, see Rönnedal (2012b).

Syntax

Alphabet. (i) A denumerably infinite set Prop of proposition letters p, q, r, s, t, p_1, q_1, r_1, s_1, t_1, p_2, q_2, r_2, s_2, t_2…, (ii) the usual primitive truth-functional connectives, (iii) the modal operators $\Box$ and $\Diamond$, (iv) the deontic operators O and P, and (v) the brackets (,).

Language. The language L is the set of well-formed formulas (wffs) generated by the usual clauses for proposition letters and propositionally compound sentences, and the following clauses: (i) if A is a wff, then $\Box$A, $\Diamond$A, OA and PA are wffs, and (iii) nothing else is a wff.

Capital letters A, B, C … are used to represent arbitrary (not necessarily atomic) formulas of the object language. $\vdash$ A, says that A is a theorem (in some system determined by the context). Outer brackets around sentences are usually dropped if the result is not ambiguous.

Definitions. $\ominus$A = $\Box\neg$A, FA = O$\neg$A, KA = PA $\wedge$ P$\neg$A, and NA = (OA $\vee$ O$\neg$A). $\bot$ (falsum) and $\top$ (verum) are defined as usual.

The translationfunction t. To understand the intended interpretation of the formal language in this essay we can use the following translation function. t($\neg$A) = It is not the case that t(A). t(A $\rightarrow$ B) = If t(A), then t(B). And similarly for all other propositional connectives. t($\Box$A) = It is necessary that t(A). t($\Diamond$A) = It is possible that t(A). t($\ominus$A) = It is impossible that t(A).

t(OA) = It ought to be the case that t(A). t(PA) = It is permitted that t(A). t(KA) = It is optional (deontically contingent) that t(A). t(NA) = It is non-optional (deontically non-contingent) that t(A). If t(p) is a sentence in English, we can use t to translate a formal sentence whose only atomic proposition letter is p into English. For instance, let t(p) be "You give money to every poor person in the whole world". Then the t-translation of "$\neg\Diamond p \rightarrow \neg Op$" is "If it is not the case that it is possible that you give money to every poor person in the whole world, then it is not the case that it ought to be the case that you give money to every poor person in the whole world". So, "$\neg\Diamond p \rightarrow \neg Op$" is the "contraposition" of one version of the so-called "ought implies can principle".

There seem to be several different kinds of necessity and possibility: logical, metaphysical, natural, historical etc. It might be plausible to use different logical systems to symbolise these different kinds. However, we can use the same symbols in each case.

Semantics

We use the same kind of semantics as in Rönnedal (2012). The only difference is that we treat $\diamondminus$, F, K and N as defined operators in this essay. The fundamental concepts are the same, the truth-conditions for various sentences are the same, the classifications of different systems are the same.

Proof theory

We will use two kinds of proof theories in this essay: one axiomatic and one that is based on semantic tableaux. Both are described in Rönnedal (2012). All fundamental axioms and tableau rules that are used in our proofs in the present paper are also described in that essay. All other rules are easily derived from the axioms and the primitive rules together with the definitions. In Rönnedal (2012) we called some axioms a-axioms (a as in "alethic), some b-axioms, and some ab-axioms. We will call the b-axioms "d-axioms" in this essay (d as in "deontic"), and the ab-axioms "ad-axioms". And similarly for the tableau rules.

3. Redundant operators

Theorem 1. The deontic operators O and P are redundant when prefixed to the alethic operators $\Box$ and $\Diamond$ in **(i)** every tableau system that contains T-a4, T-a5, T-dD and T-MO (as primitive or derived rules), and in **(ii)** every axiomatic system that contains a4, a5, dD and MO (as axioms or theorems). I.e. $\vdash \otimes A \leftrightarrow {*}\otimes A$, where $\otimes = \Box$ or $\Diamond$ and $* =$ O or P, holds in the indicated systems.

Proof. **(i)** To prove this we must show that all of the following sentences are theorems in every tableau system that includes the tableau rules T-a4, Ta5, T-dD and T-MO: $\Box A \leftrightarrow O\Box A$ ($O\Box R$), $\Box A \leftrightarrow P\Box A$ ($P\Box R$), $\Diamond A \leftrightarrow O\Diamond A$ ($O\Diamond R$), $\Diamond A \leftrightarrow P\Diamond A$ ($P\Diamond R$). We begin by proving $O\Box R$. $O\Box R$ states that it is necessary that A if and only if it is obligatory that it is necessary that A.

$O\Box R$ $\quad \Box A \leftrightarrow O\Box A$

(1) $\neg(\Box A \leftrightarrow O\Box A), 0$

$\swarrow \quad \searrow$

(2) $\Box A, 0$ [1, $\neg\leftrightarrow$]	(14) $\neg\Box A, 0$ [1, $\neg\leftrightarrow$]
(3) $\neg O\Box A, 0$ [1, $\neg\leftrightarrow$]	(15) $O\Box A, 0$ [1, $\neg\leftrightarrow$]
(4) $P\neg\Box A, 0$ [3, $\neg O$]	(16) $\Diamond\neg A, 0$ [14, $\neg\Box$]
(5) 0s1 [4, P]	(17) 0r1 [16, $\Diamond$]
(6) $\neg\Box A, 1$ [4, P]	(18) $\neg A, 1$ [16, $\Diamond$]
(7) $\Diamond\neg A, 1$ [6, $\neg\Box$]	(19) 0s2 [T-dD]
(8) 1r2 [7, $\Diamond$]	(20) $\Box A, 2$ [15, 19, O]
(9) $\neg A, 2$ [7, $\Diamond$]	(21) 0r2 [19, T-MO]
(10) 0r1 [5, T-MO]	(22) 2r1 [17, 21, T-a5]
(11) 0r2 [8, 10, T-a4]	(23) A, 1 [20, 22, $\Box$]
(12) A, 2 [2, 11, $\Box$]	(24) * [18, 23]
(13) * [9, 12]	

Hence, we see that O may be deleted when prefixed to a sentence of the form $\Box A$. Next we turn to $P\Box R$.

$P\Box R$ $\quad \Box A \leftrightarrow P\Box A$

(1) $\neg(\Box A \leftrightarrow P\Box A), 0$

$\swarrow \quad \searrow$

(2) $\Box A, 0$ [1, $\neg\leftrightarrow$]	(14) $\neg\Box A, 0$ [1, $\neg\leftrightarrow$]
(3) $\neg P\Box A, 0$ [1, $\neg\leftrightarrow$]	(15) $P\Box A, 0$ [1, $\neg\leftrightarrow$]
(4) $O\neg\Box A, 0$ [3, $\neg P$]	(16) $\Diamond\neg A, 0$ [14, $\neg\Box$]
(5) 0s1 [T-dD]	(17) 0s1 [15, P]
(6) $\neg\Box A, 1$ [4, 5, O]	(18) $\Box A, 1$ [15, P]
(7) $\Diamond\neg A, 1$ [6, $\neg\Box$]	(19) 0r2 [16, $\Diamond$]
(8) 1r2 [7, $\Diamond$]	(20) $\neg A, 2$ [16, $\Diamond$]
(9) $\neg A, 2$ [7, $\Diamond$]	(21) 0r1 [17, T-MO]
(10) 0r1 [5, T-MO]	(22) 1r2 [19, 21, T-a5]
(11) 0r2 [8, 10, T-a4]	(23) A, 2 [18, 22, $\Box$]
(12) A, 2 [2, 11, $\Box$]	(24) * [20, 23]
(13) * [9, 12]	

According to $P\Box R$, it is necessary that A if and only if it is permitted that it is necessary that A. Consequently, P may be deleted when prefixed to a sentence of the form $\Box A$.

The following tableau establishes that O◇R is a theorem in the indicated systems.

O◇R   ◇A ↔ O◇A

(1) ¬(◇A ↔ O◇A), 0

↙ ↘

(2) ◇A, 0 [1, ¬↔]	(14) ¬◇A, 0 [1, ¬↔]
(3) ¬O◇A, 0 [1, ¬↔]	(15) O◇A, 0 [1, ¬↔]
(4) P¬◇A, 0 [3, ¬O]	(16) □¬A, 0 [14, ¬◇]
(5) 0s1 [4, P]	(17) 0s1 [T-dD]
(6) ¬◇A, 1 [4, P]	(18) ◇A, 1 [15, 17, O]
(7) □¬A, 1 [6, ¬◇]	(19) 1r2 [18, ◇]
(8) 0r2 [2, ◇]	(20) A, 2 [18, ◇]
(9) A, 2 [2, ◇]	(21) 0r1 [17, T-MO]
(10) 0r1 [5, T-MO]	(22) 0r2 [19, 21, T-a4]
(11) 1r2 [8, 10, T-a5]	(23) ¬A, 2 [16, 22, □]
(12) ¬A, 2 [7, 11, □]	(24) * [20, 23]
(13) * [9, 12]	

According to O◇R, it is possible that A if and only if it is obligatory that it is possible that A. O may thus be deleted when prefixed to a sentence of the form ◇A. Finally, we prove P◇R.

P◇R   ◇A ↔ P◇A

(1) ¬(◇A ↔ P◇A), 0

↙ ↘

(2) ◇A, 0 [1, ¬↔]	(14) ¬◇A, 0 [1, ¬↔]
(3) ¬P◇A, 0 1, ¬↔]	(15) P◇A, 0 [1, ¬↔]
(4) O¬◇A, 0 [3, ¬P]	(16) □¬A, 0 [14, ¬◇]
(5) 0r1 [2, ◇]	(17) 0s1 [15, P]
(6) A, 1 [2, ◇]	(18) ◇A, 1 [15, P]
(7) 0s2 [T-dD]	(19) 1r2 [18, ◇]
(8) ¬◇A, 2 [4, 7, O]	(20) A, 2 [18, ◇]
(9) □¬A, 2 [8, ¬◇]	(21) 0r1 [17, T-MO]
(10) 0r2 [7, T-MO]	(22) 0r2 [19, 21, T-a4]
(11) 2r1 [5, 10, T-a5]	(23) ¬A, 2 [16, 22, □]
(12) ¬A, 1 [9, 11, □]	(24) * [20, 23]
(13) * [6, 12]	

P◇R says that it is possible that A if and only if it is permitted that it is possible that A. We conclude that P may be deleted when prefixed to a sentence of the form ◇A in the systems we have mentioned.

(ii) follows immediately from (i) and the soundness and completeness theorems found in Rönnedal (2012) and (2012b). However, in section 6 we will also consider some explicit axiomatic arguments. ■

The proof of theorem 1 is now finished. Due to this theorem, we may always delete the deontic operators O and P when they are prefixed to $\Box$ or $\Diamond$ in any part of any formula in the indicated systems. O$\Box$R, P$\Box$R etc. can be viewed as a kind of reduction principles similar to other well known principles of this sort.

What happens if we add F, K and N to our language? Does any of the following "reduction laws" hold: $\Box A \leftrightarrow F\Box A$, $\Diamond A \leftrightarrow F\Diamond A$, $\Box A \leftrightarrow K\Box A$, $\Diamond A \leftrightarrow K\Diamond A$, $\Box A \leftrightarrow N\Box A$, $\Diamond A \leftrightarrow N\Diamond A$? The answer to this question is no. However we have: $F\Box A \leftrightarrow \neg\Box A$, $F\Diamond A \leftrightarrow \neg\Diamond A$, $K\Box A \leftrightarrow \bot$, $K\Diamond A \leftrightarrow \bot$, $N\Box A \leftrightarrow \top$, and $N\Diamond A \leftrightarrow \top$. So, in one sense F, K and N are redundant when prefixed to $\Box$ or $\Diamond$. However, if we delete them from a formula this might affect the formula's truth-value.

4. Necessarily equivalent sentences and deontic status

Theorem 2. Necessarily equivalent sentences have the same deontic status with respect to O, P, F, K and N in **(i)** every tableau system that contains T-MO, and in **(ii)** every axiomatic system that contains MO. More precisely: $\vdash \Box(A \leftrightarrow B) \rightarrow ({*A} \leftrightarrow {*B})$, where * = O, P, F, K and N, hold in the indicated systems. If two sentences do not have the same deontic status with respect to O, P, F, K and N in a tableau system that contains T-MO or in an axiomatic system that contains MO, then they are not necessarily equivalent. More precisely: $\vdash \neg({*A} \leftrightarrow {*B}) \rightarrow \neg\Box(A \leftrightarrow B)$, where * = O, P, F, K and N, holds in **(iii)** every tableau system that contains T-MO, and in **(iv)** every axiomatic system that contains MO.[1]

Proof. **(i)** To prove this theorem we have to show that all of the following sentences are theorems in every tableau system that contains T-MO: $\Box(A \leftrightarrow B) \rightarrow (PA \leftrightarrow PB)$ (PE), $\Box(A \leftrightarrow B) \rightarrow (OA \leftrightarrow OB)$ (OE), $\Box(A \leftrightarrow B) \rightarrow (FA \leftrightarrow FB)$ (FE), $\Box(A \leftrightarrow B) \rightarrow (KA \leftrightarrow KB)$ (KE), and $\Box(A \leftrightarrow B) \rightarrow (NA \leftrightarrow NB)$ (NE). The tableau proofs of (OE) and (NE) are left to the reader. However, we will see how it is easy (by an axiomatic argument) to establish (NE) once we have proven (KE). All we have to do then is to prove (PE), (FE), and (KE). Let us begin with (PE).

(PE) says that if it is necessary that t(A) if and only if t(B), then it is permitted that t(A) if and only if it is permitted that t(B), according to the t-translation of this sentence.

[1] Of course, $\Box(A \leftrightarrow B) \rightarrow (\Box A \leftrightarrow \Box B)$ and $\Box(A \leftrightarrow B) \rightarrow (\Diamond A \leftrightarrow \Diamond B)$ are also theorems in every alethic-deontic system described in Rönnedal (2012).

$\Box(A \leftrightarrow B) \rightarrow (PA \leftrightarrow PB)$

$(1)\ \Box(A \leftrightarrow B) \rightarrow (PA \leftrightarrow PB), 0$
$(2)\ \Box(A \leftrightarrow B), 0$
$(3)\ \neg(PA \leftrightarrow PB), 0$
$\swarrow \quad \searrow$

$(4)\ PA, 0$	$(18)\ \neg PA, 0$
$(5)\ \neg PB, 0$	$(19)\ PB, 0$
$(6)\ O\neg B, 0$	$(20)\ O\neg A, 0$
$(7)\ 0s1$	$(21)\ 0s1$
$(8)\ A, 1$	$(22)\ B, 1$
$(9)\ \neg B, 1$	$(23)\ \neg A, 1$
$(10)\ 0r1$	$(24)\ 0r1$
$(11)\ A \leftrightarrow B, 1$	$(25)\ A \leftrightarrow B, 1$
$\swarrow \quad \searrow$	$\swarrow \quad \searrow$

$(12)\ A, 1$	$(15)\ \neg A, 1$	$(26)\ A, 1$	$(29)\ \neg A, 1$
$(13)\ B, 1$	$(16)\ \neg B, 1$	$(27)\ B, 1$	$(30)\ \neg B, 1$
(14) *	(17) *	(28) *	(31) *

Next we continue with (FE). The t-translation of (FE) looks like this: If it is necessary that t(A) if and only if t(B), then it is forbidden that t(A) if and only if it is forbidden that t(B).

$\Box(A \leftrightarrow B) \rightarrow (FA \leftrightarrow FB)$

$(1)\ \neg(\Box(A \leftrightarrow B) \rightarrow (FA \leftrightarrow FB)), 0$
$(2)\ \Box(A \leftrightarrow B), 0$
$(3)\ \neg(FA \leftrightarrow FB), 0$
$\swarrow \quad \searrow$

$(4)\ FA, 0$	$(19)\ \neg FA, 0$
$(5)\ \neg FB, 0$	$(20)\ FB, 0$
$(6)\ O\neg A, 0$	$(21)\ O\neg B, 0$
$(7)\ PB, 0$	$(22)\ PA, 0$
$(8)\ 0s1$	$(23)\ 0s1$
$(9)\ B, 1$	$(24)\ A, 1$
$(10)\ \neg A, 1$	$(25)\ \neg B, 1$
$(11)\ 0r1$	$(26)\ 0r1$
$(12)\ A \leftrightarrow B, 1$	$(27)\ A \leftrightarrow B, 1$
$\swarrow \quad \searrow$	$\swarrow \quad \searrow$

$(13)\ A, 1$	$(16)\ \neg A, 1$	$(28)\ A, 1$	$(31)\ \neg A, 1$
$(14)\ B, 1$	$(17)\ \neg B, 1$	$(29)\ B, 1$	$(32)\ \neg B, 1$
(15) *	(18) *	(30) *	(33) *

Since we have not introduced any special rules for deontic contingency and non-contingency, or optionality and non-optionality, we must first translate the sentence (KE) to be able to decide whether it is valid or not by use of the tableau method. By definition, $\Box(A \leftrightarrow B) \rightarrow (KA \leftrightarrow KB)$ is equivalent to $\Box(A \leftrightarrow B) \rightarrow ((PA \wedge P\neg A) \leftrightarrow (PB \wedge P\neg B))$. To prove that the former is a theorem in our indicated systems, it thus suffices to prove the latter.

If we apply the t-translation function to (KE), it says that if it is necessary that t(A) if and only if t(B), then it is optional that t(A) if and only if it is optional that t(B). Here is the tableau proof.

$\Box(A \leftrightarrow B) \rightarrow (KA \leftrightarrow KB)$

```
              (1) ¬(□(A ↔ B) → ((PA ∧ P¬A) ↔ (PB ∧ P¬B))), 0
                              (2) □(A ↔ B), 0
                     (3) ¬((PA ∧ P¬A) ↔ (PB ∧ P¬B)), 0
                                  ↙  ↘
              (4) PA ∧ P¬A, 0            (34) ¬(PA ∧ P¬A), 0
              (5) ¬(PB ∧ P¬B), 0         (35) PB ∧ P¬B, 0
              (6) PA, 0                  (36) PB, 0
              (7) P¬A, 0                 (37) P¬B, 0
              (8) 0s1                    (38) 0s1
              (9) A, 1                   (39) B, 1
              (10) 0s2                   (40) 0s2
              (11) ¬A, 2                 (41) ¬B, 2
              (12) 0r1                   (42) 0r1
              (13) 0r2                   (43) 0r2
              (14) A ↔ B, 1              (44) A ↔ B, 1
              (15) A ↔ B, 2              (45) A ↔ B, 2
                 ↙      ↘                   ↙      ↘
        (16) A, 1     (31) ¬A, 1    (46) A, 1     (61) ¬A, 1
        (17) B, 1     (32) ¬B, 1    (47) B, 1     (62) ¬B, 1
        ↙       ↘     (33) *        ↙       ↘     (63) *
(18) A, 2     (21) ¬A, 2    (48) A, 2     (51) ¬A, 2
(19) B, 2     (22) ¬B, 2    (49) B, 2     (52) ¬B, 2
(20) *        ↙       ↘     (50) *          ↙       ↘
        (23) ¬PB, 0   (27) ¬P¬B, 0      (53) ¬PA, 0   (57) ¬P¬A, 0
        (24) O¬B, 0   (28) O¬¬B, 0      (54) O¬A, 0   (58) O¬¬A, 0
        (25) ¬B, 1    (29) ¬¬B, 2       (55) ¬A, 1    (59) ¬¬A, 2
        (26) *        (30) *            (56) *        (60) *
```

I will omit the tableau proof of (NE). It is similar to the one just given. Instead we establish the result by the following reasoning. 1. $\Box(A \leftrightarrow B) \rightarrow$

$(KA \leftrightarrow KB)$ [From the tableau proof above]. 2. $\Box(A \leftrightarrow B) \rightarrow (\neg KA \leftrightarrow \neg KB)$ [1, propositional logic]. 3. $\Box(A \leftrightarrow B) \rightarrow (NA \leftrightarrow NB)$ [by 2 and the definitions of K and N]. We know that this kind of reasoning (and the kind exhibited in the proof of KE) is correct, since we have proved that all our axiomatic systems as well as our tableau systems are sound and complete with respect to the same semantics (Rönnedal (2012), (2012b)).

(ii) follows immediately from (i) and the soundness and completeness theorems found in Rönnedal (2012) and (2012b).

(iii) and **(iv).** This is essentially the contrapositive of part (i) and part (ii), respectively. (Details are easy and are left to the reader.) ■

5. A partition of the sentences in T-dD and dD systems

Theorem 3. The set of all sentences (in our formal language) can be partitioned into the following, mutually exclusive, exhaustive subsets in **(i)** every tableau system that includes T-dD (as a primitive or derived rule), and in **(ii)** every axiomatic system that includes dD (as an axiom or theorem).

$\Box A \wedge OA$	$OA \wedge \neg\Box A$	$PA \wedge P\neg A$	$FA \wedge \neg\Diamond A$	$FA \wedge \Diamond A$

Proof. **(i)** To prove this theorem we must show that every sentence is contained in at least one of these categories and that it is contained in at most one of them. This amounts to proving the following theorems.

$P = (\Box A \wedge OA) \vee (OA \wedge \neg\Box A)$
$\quad \vee (PA \wedge P\neg A) \vee (FA \wedge \neg\Diamond A) \vee (FA \wedge \Diamond A).$

$A = (\Box A \wedge OA) \rightarrow$
$\quad (\neg(OA \wedge \neg\Box A) \wedge \neg(PA \wedge P\neg A) \wedge \neg(FA \wedge \neg\Diamond A) \wedge \neg(FA \wedge \Diamond A)).$

$B = (OA \wedge \neg\Box A) \rightarrow$
$\quad (\neg(\Box A \wedge OA) \wedge \neg(PA \wedge P\neg A) \wedge \neg(FA \wedge \neg\Diamond A) \wedge \neg(FA \wedge \Diamond A)).$

$C = (PA \wedge P\neg A) \rightarrow$
$\quad (\neg(OA \wedge \neg\Box A) \wedge \neg(\Box A \wedge OA) \wedge \neg(FA \wedge \neg\Diamond A) \wedge \neg(FA \wedge \Diamond A)).$

$D = (FA \wedge \neg\Diamond A) \rightarrow$
$\quad (\neg(OA \wedge \neg\Box A) \wedge \neg(PA \wedge P\neg A) \wedge \neg(\Box A \wedge OA) \wedge \neg(FA \wedge \Diamond A)).$

$E = (FA \wedge \Diamond A) \rightarrow$
$\quad (\neg(OA \wedge \neg\Box A) \wedge \neg(PA \wedge P\neg A) \wedge \neg(FA \wedge \neg\Diamond A) \wedge \neg(\Box A \wedge OA)).$

A is equivalent to the conjunction of the following sentences:

A1 $(\Box A \wedge OA) \rightarrow \neg(OA \wedge \neg\Box A)$, A2 $(\Box A \wedge OA) \rightarrow \neg(PA \wedge P\neg A)$,
A3 $(\Box A \wedge OA) \rightarrow \neg(FA \wedge \neg\Diamond A)$, A4 $(\Box A \wedge OA) \rightarrow \neg(FA \wedge \Diamond A)$.

B is equivalent to the conjunction of the following sentences:

B1 $(OA \wedge \neg\Box A) \rightarrow \neg(\Box A \wedge OA)$, B2 $(OA \wedge \neg\Box A) \rightarrow \neg(PA \wedge P\neg A)$, B3 $(OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \neg\diamondminus A)$, B4 $(OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \diamondminus A)$.

C is equivalent to the conjunction of the following sentences: C1 $(PA \wedge P\neg A) \rightarrow \neg(OA \wedge \neg\Box A)$, C2 $(PA \wedge P\neg A) \rightarrow \neg(\Box A \wedge OA)$, C3 $(PA \wedge P\neg A) \rightarrow \neg(FA \wedge \neg\diamondminus A)$, C4 $(PA \wedge P\neg A) \rightarrow \neg(FA \wedge \diamondminus A)$.

D is equivalent to the conjunction of the following sentences: D1 $(FA \wedge \neg\diamondminus A) \rightarrow \neg(OA \wedge \neg\Box A)$, D2 $(FA \wedge \neg\diamondminus A) \rightarrow \neg(PA \wedge P\neg A)$, D3 $(FA \wedge \neg\diamondminus A) \rightarrow \neg(\Box A \wedge OA)$, D4 $(FA \wedge \neg\diamondminus A) \rightarrow \neg(FA \wedge \diamondminus A)$.

E is equivalent to the conjunction of the following sentences: E1 $(FA \wedge \diamondminus A) \rightarrow \neg(OA \wedge \neg\Box A)$, E2 $(FA \wedge \diamondminus A) \rightarrow \neg(PA \wedge P\neg A)$, E3 $(FA \wedge \diamondminus A) \rightarrow \neg(FA \wedge \neg\diamondminus A)$, E4 $(FA \wedge \diamondminus A) \rightarrow \neg(\Box A \wedge OA)$.

So, it is enough that we show that P, A1-A4, B1-B4, C1-C4, D1-D4 and E1-E4 are theorems in our systems. A1, B1, D4 and E3 are true by propositional logic alone. We continue to prove A3 and A4.

A3 $(\Box A \wedge OA) \rightarrow \neg(FA \wedge \neg\diamondminus A)$

(1) $\neg((\Box A \wedge OA) \rightarrow \neg(FA \wedge \neg\diamondminus A))$, 0
(2) $\Box A \wedge OA$, 0 [1, $\neg\rightarrow$]
(3) $\neg\neg(FA \wedge \neg\diamondminus A)$, 0 [1, $\neg\rightarrow$]
(4) $FA \wedge \neg\diamondminus A$, 0 [3, $\neg\neg$]
(5) $\Box A$, 0 [2, $\wedge$]
(6) OA, 0 [2, $\wedge$]
(7) FA, 0 [4, $\wedge$]
(8) $\neg\diamondminus A$, 0 [4, $\wedge$]
(9) 0s1 [T-dD]
(10) A, 1 [6, 9, O]
(11) $\neg A$, 1 [7, 9, F’]
(12) * [10, 11]

A4 $(\Box A \wedge OA) \rightarrow \neg(FA \wedge \diamondminus A)$

(1) $\neg((\Box A \wedge OA) \rightarrow \neg(FA \wedge \diamondminus A))$, 0
(2) $\Box A \wedge OA$, 0 [1, $\neg\rightarrow$]
(3) $\neg\neg(FA \wedge \diamondminus A)$, 0 [1, $\neg\rightarrow$]
(4) $FA \wedge \diamondminus A$, 0 [3, $\neg\neg$]
(5) $\Box A$, 0 [2, $\wedge$]
(6) OA, 0 [2, $\wedge$]
(7) FA, 0 [4, $\wedge$]
(8) $\diamondminus A$, 0 [4, $\wedge$]
(9) 0s1 [T-dD]
(10) A, 1 [6, 9, O]
(11) $\neg A$, 1 [7, 9, F’]
(12) * [10, 11]

B3 $(OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \neg\diamondminus A)$

(1) $\neg((OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \neg\diamondminus A))$, 0
(2) $OA \wedge \neg\Box A$, 0 [1, $\neg\rightarrow$]
(3) $\neg\neg(FA \wedge \neg\diamondminus A)$, 0 [1, $\neg\rightarrow$]
(4) $FA \wedge \neg\diamondminus A$, 0 [3, $\neg\neg$]
(5) OA, 0 [2, $\wedge$]
(6) $\neg\Box A$, 0 [2, $\wedge$]
(7) FA, 0 [4, $\wedge$]
(8) $\neg\diamondminus A$, 0 [4, $\wedge$]

B4 $(OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \diamondminus A)$

(1) $\neg((OA \wedge \neg\Box A) \rightarrow \neg(FA \wedge \diamondminus A))$, 0
(2) $OA \wedge \neg\Box A$, 0 [1, $\neg\rightarrow$]
(3) $\neg\neg(FA \wedge \diamondminus A)$, 0 [1, $\neg\rightarrow$]
(4) $FA \wedge \diamondminus A$, 0 [3, $\neg\neg$]
(5) OA, 0 [2, $\wedge$]
(6) $\neg\Box A$, 0 [2, $\wedge$]
(7) FA, 0 [4, $\wedge$]
(8) $\diamondminus A$, 0 [4, $\wedge$]

(9) 0s1 [T-dD]
(10) A, 1 [5, 9, O]
(11) ¬A, 1 [7, 9, F’]
(12) * [10, 11]

(9) 0s1 [T-dD]
(10) A, 1 [5, 9, O]
(11) ¬A, 1 [7, 9, F’]
(12) * [10, 11]

D1 is equivalent to B3 and E1 is equivalent to B4. Consequently, D1 and E1 are theorems in all T-dD systems. For the tableaux above prove that both B3 and B4 are theorems in systems of this kind.

A2 (□A ∧ OA) → ¬(PA ∧ P¬A)

(1) ¬((□A ∧ OA) → ¬(PA ∧ P¬A)), 0
(2) □A ∧ OA, 0 [1, ¬→]
(3) ¬¬(PA ∧ P¬A), 0 [1, ¬→]
(4) PA ∧ P¬A, 0 [3, ¬¬]
(5) □A, 0 [2, ∧]
(6) OA, 0 [2, ∧]
(7) PA, 0 [4, ∧]
(8) P¬A, 0 [4, ∧]
(9) 0s1 [8, P]
(10) ¬A, 1 [8, P]
(11) A, 1 [6, 9, O]
(12) * [10, 11]

B2 (OA ∧ ¬□A) → ¬(PA ∧ P¬A)

(1) ¬((OA ∧ ¬□A) → ¬(PA ∧ P¬A)), 0
(2) OA ∧ ¬□A, 0 [1, ¬→]
(3) ¬¬(PA ∧ P¬A), 0 [1, ¬→]
(4) PA ∧ P¬A, 0 [3, ¬¬]
(5) OA, 0 [2, ∧]
(6) ¬□A, 0 [2, ∧]
(7) PA, 0 [4, ∧]
(8) P¬A, 0 [4, ∧]
(9) 0s1 [8, P]
(10) ¬A, 1 [8, P]
(11) A, 1 [5, 9, O]
(12) * [10, 11]

We now know that A2 and B2 are theorems in all T-dD systems. C2 is the contrapositive of A2 and C1 is the contrapositive of B2. It follows that also C2 and C1 are theorems in all systems of this kind.

D2 (FA ∧ ¬◇A) → ¬(PA ∧ P¬A)

(1) ¬((FA ∧ ¬◇A) → ¬(PA ∧ P¬A)), 0
(2) FA ∧ ¬◇A, 0 [1, ¬→]
(3) ¬¬(PA ∧ P¬A), 0 [1, ¬→]
(4) PA ∧ P¬A, 0 [3, ¬¬]
(5) FA, 0 [2, ∧]
(6) ¬◇A, 0 [2, ∧]
(7) PA, 0 [4, ∧]
(8) P¬A, 0 [4, ∧]
(9) 0s1 [7, P]
(10) A, 1 [7, P]
(11) ¬A, 1 [5, 9, F’]
(12) * [10, 11]

E2 (FA ∧ ◇A) → ¬(PA ∧ P¬A)

(1) ¬((FA ∧ ◇A) → ¬(PA ∧ P¬A)), 0
(2) FA ∧ ◇A, 0 [1, ¬→]
(3) ¬¬(PA ∧ P¬A), 0 [1, ¬→]
(4) PA ∧ P¬A, 0 [3, ¬¬]
(5) FA, 0 [2, ∧]
(6) ◇A, 0 [2, ∧]
(7) PA, 0 [4, ∧]
(8) P¬A, 0 [4, ∧]
(9) 0s1 [7, P]
(10) A, 1 [7, P]
(11) ¬A, 1 [5, 9, F’]
(12) * [10, 11]

These tableaux prove that D2 and E2 are theorems in all T-dD systems. D2 is the contrapositive of C3 and E2 is the contrapositive of C4. Hence, C3 and C4 are also theorems in every T-dD system. D3 is the contrapositive of A3 and E4 is the contrapositive of A4. It follows that D3 and E3 are theorems in all T-dD systems.

We have now established A, B, C, D and E. All that is left to prove is that (P) is a theorem in all tableau systems that include T-dD.

Consider (P) $(\Box A \wedge OA) \vee (OA \wedge \neg\Box A) \vee (PA \wedge P\neg A) \vee (FA \wedge \neg\Diamond A) \vee (FA \wedge \Diamond A)$. We can prove this sentence directly by constructing a semantic tableau that starts with the negation of (P). But the proof is quite long. So, I will combine axiomatic and tableau techniques. First we show that $OA \vee (PA \wedge P\neg A) \vee FA$ is a theorem in every system that includes T-dD. Indeed, this sentence is a theorem even in the weakest so-called normal deontic logic, sometimes called OK. The theorem states that everything is obligatory, optional or forbidden.

P $\quad OA \vee ((PA \wedge P\neg A) \vee FA)$

(1) $\neg(OA \vee ((PA \wedge P\neg A) \vee FA))$, 0
(2) $\neg OA$, 0 [1, $\neg\vee$]
(3) $\neg((PA \wedge P\neg A) \vee FA)$, 0 [1, $\neg\vee$]
(4) $\neg(PA \wedge P\neg A)$, 0 [3, $\neg\vee$]
(5) $\neg FA$, 0 [3, $\neg\vee$]
(6) $P\neg A$, 0 [2, $\neg O$]
(7) PA, 0 [5, $\neg F$]
(8) 0s1 [7, P]
(9) A, 1 [7, P]
(10) 0s2 [6, P]
(11) $\neg A$, 2 [6, P]
↙ ↘

(12) $\neg PA$, 0 [4, $\neg\wedge$]	(16) $\neg P\neg A$, 0 [4, $\neg\wedge$]
(13) $O\neg A$, 0 [12, $\neg P$]	(17) $O\neg\neg A$, 0 [16, $\neg P$]
(14) $\neg A$, 1 [8, 13, O]	(18) $\neg\neg A$, 2 [10, 17, O]
(15) * [9, 14]	(19) * [11, 18]

Now we can reason as follows. We know that A is equivalent to $(A \wedge B) \vee (A \wedge \neg B)$. Hence, OA is equivalent to $(OA \wedge \Box A) \vee (OA \wedge \neg\Box A)$ and FA is equivalent to $(FA \wedge \neg\Diamond A) \vee (FA \wedge \Diamond A)$. So, we substitute OA by $(OA \wedge \Box A) \vee (OA \wedge \neg\Box A)$ and FA by $(FA \wedge \neg\Diamond A) \vee (FA \wedge \Diamond A)$ in $OA \vee (PA \wedge P\neg A) \vee FA$. It follows that $(\Box A \wedge OA) \vee (OA \wedge \neg\Box A) \vee (PA \wedge P\neg A) \vee (FA \wedge \neg\Diamond A) \vee (FA \wedge \Diamond A)$ is a theorem. Our proof of theorem 3 is now finished. ■

6. Modalities in some systems

Theorem 4. In theorem 4 we use a language without the defined concepts in section 2. **(i)** In every axiomatic (normal) alethic-deontic logic that contains the deontic system OS5+, the alethic system S5[2], MO, OC, ad4 $OA \to \Box OA$, and ad5 $PA \to \Box PA$ (as axioms and/or theorems) there are at most ten distinct modalities: A, $\neg A$, $\Diamond A$, $\Box A$, PA, OA, $\neg\Diamond A$, $\neg\Box A$, $\neg PA$ and $\neg OA$. **(ii)** In the axiomatic alethic-deontic system aS5dOS5+adMOOC45[3] there are exactly ten distinct modalities: A, $\neg A$, $\Diamond A$, $\Box A$, PA, OA, $\neg\Diamond A$, $\neg\Box A$, $\neg PA$ and $\neg OA$. **(iii)** In every normal alethic-deontic tableau system that includes T-aT, T-aB, T-a4, T-dD, T-d4, T-d5, T-MO, T-OC, T-ad4, and T-ad5 (as primitive and/or derived rules) there are at most ten distinct modalities: A, $\neg A$, $\Diamond A$, $\Box A$, PA, OA, $\neg\Diamond A$, $\neg\Box A$, $\neg PA$ and $\neg OA$. **(iv)** In the alethic-deontic tableau system T-aTB4dD45adMOOC45[4] there are exactly ten distinct modalities: A, $\neg A$, $\Diamond A$, $\Box A$, PA, OA, $\neg\Diamond A$, $\neg\Box A$, $\neg PA$ and $\neg OA$. **(v)** In the axiomatic alethic-deontic system aS5dOS5+adMOOC45 and in the alethic-deontic tableau system T-aTB4dD45adMOOC45 (and in all systems that are deductively equivalent) a string of modal operators reduces to its innermost modality. E.g. all of the following equivalences hold in these systems: $OO\Diamond A \leftrightarrow \Diamond A$, $\neg\Box\Box OPA \leftrightarrow \neg PA$, $\Box\Box\neg\Diamond POOA \leftrightarrow \neg OA$, $P\Box\Diamond\Diamond\neg P\neg PA \leftrightarrow PA$, $\Diamond\Box OP\Box A \leftrightarrow \Box A$, $\neg\Diamond\Box\Box\neg OA \leftrightarrow OA$, $\neg P\neg\Box\Box\Box\neg\Diamond A \leftrightarrow \neg\Diamond A$, $\neg O\neg\Diamond\Diamond\Diamond\Box O\neg\Box A \leftrightarrow \neg\Box A$.

Proof. **(i)** The proof of this theorem consists of two parts. First we show that the systems we are interested in contain the following reduction laws:

> (i) $A \leftrightarrow \neg\neg A$, (ii) $\Diamond A \leftrightarrow \Diamond\Diamond A$, (iii) $\Box A \leftrightarrow \Box\Box A$, (iv) $\Diamond A \leftrightarrow \Box\Diamond A$, (v) $\Box A \leftrightarrow \Diamond\Box A$, (vi) $PA \leftrightarrow PPA$, (vii) $OA \leftrightarrow OOA$, (viii) $PA \leftrightarrow OPA$, (ix) $OA \leftrightarrow POA$, (x) $\Diamond A \leftrightarrow P\Diamond A$, (xi) $\Diamond A \leftrightarrow O\Diamond A$, (xii) $\Box A \leftrightarrow P\Box A$, (xiii) $\Box A \leftrightarrow O\Box A$, (xiv) $PA \leftrightarrow \Diamond PA$, (xv) $PA \leftrightarrow \Box PA$, (xvi) $OA \leftrightarrow \Diamond OA$, (xvii) $OA \leftrightarrow \Box OA$.

Then we prove that given these reduction laws any modality is equivalent to one of the ten modalities mentioned above. To prove the first step, we use axiomatic techniques. It is a well known fact that (ii)-(v) hold in S5 and it has been established that (vi)-(ix) are theorems in OS5+.[5] All that remains to show then is (i) and (x)-(xvii). (i) is simply the law of double negation and (xii)-(xvii) can be obtained in the following manner.

---

[2] See e.g. Rönnedal (2010), chapter 7, for more information about OS5+ (also called KD45). S5 is described in almost every introduction to (alethic) modal logic.

[3] This system contains some redundancy and there are a number a different deductively equivalent systems. Therefore, the conclusion also holds in many “other” systems.

[4] See the comment in footnote 3.

[5] See e.g. Rönnedal (2010), pp. 260-265.

(x) $\Diamond A \leftrightarrow P\Diamond A$

1. $\Box\Diamond A \rightarrow P\Diamond A$	[OC’ $\Diamond A/A$]
2. $\Diamond A \rightarrow \Box\Diamond A$	[S5]
3. $\Diamond A \rightarrow P\Diamond A$	[1, 2, PL]
4. $P\Diamond A \rightarrow \Diamond\Diamond A$	[MO’ $\Diamond A/A$]
5. $\Diamond\Diamond A \rightarrow \Diamond A$	[S5]
6. $P\Diamond A \rightarrow \Diamond A$	[4, 5, PL]
7. $\Diamond A \leftrightarrow P\Diamond A$	[3, 6, PL]

---

1. $P\Box A \rightarrow \Diamond\Box A$	[MO’ $\Box A/A$]

(xi) $\Diamond A \leftrightarrow O\Diamond A$

2. $\Diamond\Box A \rightarrow \Box A$	[S5]
3. $P\Box A \rightarrow \Box A$	[1, 2, PL]
4. $\neg\Box A \rightarrow \neg P\Box A$	[3, PL]
5. $\neg\Box\neg A \rightarrow \neg P\Box\neg A$	[4, $\neg A/A$]
6. $\Diamond A \rightarrow \neg P\neg\Diamond A$	[5, $\Box\Diamond$I]
7. $\Diamond A \rightarrow O\Diamond A$	[6, OPI]
8. $O\Diamond A \rightarrow \Diamond\Diamond A$	[OC $\Diamond A/A$]
9. $\Diamond\Diamond A \rightarrow \Diamond A$	[S5]
10. $O\Diamond A \rightarrow \Diamond A$	[8, 9, PL]
11. $\Diamond A \leftrightarrow O\Diamond A$	[7, 10, PL]

Theorem 4**(i)**, part (x)-(xi)

(xii) $\Box A \leftrightarrow P\Box A$

1. $\Box\Box A \rightarrow P\Box A$	[OC’ $\Box A/A$]
2. $\Box A \rightarrow \Box\Box A$	[S5]
3. $\Box A \rightarrow P\Box A$	[1, 2, PL]
4. $P\Box A \rightarrow \Diamond\Box A$	[MO’ $\Box A/A$]
5. $\Diamond\Box A \rightarrow \Box A$	[S5]
6. $P\Box A \rightarrow \Box A$	[4, 5, PL]
7. $\Box A \leftrightarrow P\Box A$	[3, 6, PL]

(xiii) $\Box A \leftrightarrow O\Box A$

1. $\Box\Box A \rightarrow O\Box A$	[MO $\Box A/A$]
2. $\Box A \rightarrow \Box\Box A$	[S5]
3. $\Box A \rightarrow O\Box A$	[1, 2, PL]
4. $O\Box A \rightarrow \Diamond\Box A$	[OC $\Box A/A$]
5. $\Diamond\Box A \rightarrow \Box A$	[S5]
6. $O\Box A \rightarrow \Box A$	[4, 5, PL]
7. $\Box A \leftrightarrow O\Box A$	[3, 6, PL]

Theorem 4**(i)**, part (xii)-(xiii)

(xiv) $PA \leftrightarrow \Diamond PA$

1. $PA \rightarrow \Diamond PA$	[aT’ PA/A]
2. $OA \rightarrow \Box OA$	[ad4]
3. $\neg\Box OA \rightarrow \neg OA$	[2, PL]
4. $\neg\Box O\neg A$ $\rightarrow \neg O\neg A$	[3, $\neg A/A$]
5. $\neg\Box\neg PA \rightarrow PA$	[4, OPI]
6. $\Diamond PA \rightarrow PA$	[5, $\Box\Diamond$I]
7. $PA \leftrightarrow \Diamond PA$	[1, 6, PL]

(xvi) $OA \leftrightarrow \Diamond OA$

1. $OA \rightarrow \Diamond OA$	[aT’ OA/A]
2. $PA \rightarrow \Box PA$	[ad5]
3. $\neg\Box PA \rightarrow \neg PA$	[2, PL]
4. $\neg\Box P\neg A$ $\rightarrow \neg P\neg A$	[3, $\neg A/A$]
5. $\neg\Box\neg OA \rightarrow OA$	[4, OPI]
6. $\Diamond OA \rightarrow OA$	[5, $\Box\Diamond$I]
7. $OA \leftrightarrow \Diamond OA$	[1, 6, PL]

Theorem 4**(i)**, part (xiv), (xvi)

Note that $\Diamond PA \rightarrow PA$ is the dual of $OA \rightarrow \Box OA$ and that $\Diamond OA \rightarrow OA$ is the dual of $PA \rightarrow \Box PA$.

(xv) $PA \leftrightarrow \Box PA$		(xvii) $OA \leftrightarrow \Box OA$	
1. $PA \rightarrow \Box PA$	[ad5]	1. $OA \rightarrow \Box OA$	[ad4]
2. $\Box PA \rightarrow PA$	[aT PA/A]	2. $\Box OA \rightarrow OA$	[aT OA/A]
3. $PA \leftrightarrow \Box PA$	[1, 2, PL]	3. $OA \leftrightarrow \Box OA$	[1, 2, PL]

Theorem 4**(i)**, part (xv), (xvii)

We have now proved that all of the reduction laws mentioned above hold in every system that contains S5, OS5+, MO, OC, OA → □OA, and PA → □PA.[6] It remains to show that any system that contains these laws have at most the ten distinct modalities listed in the theorem. We do this by systematically adding single modalities to the empty modality and show how the result ultimately reduces to one of the ten modalities.

So, start with the empty modality (0) *A. Add one modality. Then we get (1) ¬A, (2) ◇A, (3) □A, (4) PA, or (5) OA.

Add one modality to (1). Then we get (1.1) ¬¬A, which is equivalent to (1), (1.2) ◇¬A, which is equivalent to ¬□A, (1.3) □¬A, which is equivalent to ¬◇A, (1.4) P¬A, which is equivalent to ¬OA, or (1.5) O¬A, which is equivalent to ¬PA. Add one modality to (1.2). Then we get (1.2.1) ¬◇¬A, which is equivalent to □A (see (3)), or we get (1.2.2) ◇◇¬A, (1.2.3) □◇¬A, (1.2.4) P◇¬A, or (1.2.5) O◇¬A, all of which are equivalent to (1.2). Add one modality to 1.3. Then we get (1.3.1) ¬□¬A, which is equivalent to ◇A (see (2)), or we get (1.3.2) ◇□¬A, (1.3.3) □□¬A, (1.3.4) P□¬A, or (1.3.5) O□¬A, all of which are equivalent to (1.3). Add one modality to (1.4). Then we get (1.4.1) ¬P¬A, which is equivalent to OA (see (5)), or we get (1.4.2) ◇P¬A, (1.4.3) □P¬A, (1.4.4) PP¬A, or (1.4.5) OP¬A, all of which are equivalent to (1.4). Add one modality to (1.5). Then we get (1.5.1) ¬O¬A, which is equivalent to PA (see (2)), or we get (1.5.2) ◇O¬A, (1.5.3) □O¬A, (1.5.4) PO¬A, or (1.5.5) OO¬A, all of which are equivalent to (1.5).

Add one modality to (2). Then we get (2.1) ¬◇A (see (1.3)), (2.2) ◇◇A, (2.3) □◇A, (2.4) P◇A, or (2.5), all of which are equivalent to (2).

Add one modality to (3). Then we get (3.1) ¬□A (see (1.2)), (3.2) ◇□A, (3.3) □□A, (3.4) P□A, or (3.5) O□A, all of which are equivalent to (3).

Add one modality to (4). Then we get (4.1) ¬PA (see (1.5)), (4.2) ◇PA, (4.3) □PA, (4.4) PPA, or (4.5) OPA, all of which are equivalent to (4).

[6] In the proofs above, PL means that the step follows by propositional logic. S5 means that the sentence is a theorem in the alethic system S5. aT' is the dual of aT, MO' is the dual of MO, and OC' is the dual of OC. □◇I includes the usual relationships between □ and ◇ and OPI the usual relationships between O and P.

Add one modality to (5). Then we get (5.1) ¬OA (see (1.4)), (5.2) ◇OA, (5.3) □OA, (5.4) POA, or (5.5) OOA, all of which are equivalent to (5). This takes care of all the possibilities.

(ii) To prove (ii) we must show that the system aS5dOS5+adMOOC45 doesn't contain any more reduction laws, i.e. we must show that it is not the case that □A ↔ ◇A, OA ↔ PA, □A ↔ OA, ◇A ↔ PA etc. We can do this by describing a countermodel for each equivalence of this kind. Details are left to the reader.

(iii) Follows directly from (i) by the soundness and completeness theorems in Rönnedal (2012). See also Rönnedal (2012b).

(iv) To prove part (iv) we can use the same countermodels as in (ii).

(v) Follows from the previous parts of this theorem.

The proof of theorem 4 is now finished. ■

References

Anderson, A. R, (1956). The formal analysis of normative systems. In N. Rescher (ed.), *The Logic of Decision and Action*, Pittsburgh: University of Pittsburgh Press, 1967, pp. 147-213.

Blackburn, P., de Rijke, M. & Venema, Y. (2001). *Modal Logic*. Cambridge University Press.

Blackburn, P., van Benthem, J., Wolter, F. (eds.). (2007). *Handbook of Modal Logic*. Elsevier.

Chellas, B. F. (1980). *Modal Logic: An Introduction*. Cambridge: Cambridge University Press.

Fine, K. & Schurz, G. (1996). Transfer Theorems for Multimodal Logics, in J. Copeland (ed.) (1996), *Logic and Reality. Essays in Pure and Applied Logic. In Memory of Arthur Prior*, Oxford University Press, Oxford, pp. 169-213.

Fitting, M. & Mendelsohn, R. L. (1998). *First-Order Modal Logic*. Kluwer Academic Publishers.

Gabbay, D. M. (1976). *Investigations in Modal and Tense Logics with Applications to Problems in Philosophy and Linguistics*. Dordrecht: D. Reidel Publishing Company.

Gabbay, D., Horty, J., Parent, X., van der Meyden, E. & van der Torre, L. (eds.). (2013). *Handbook of Deontic Logic and Normative Systems*. College Publications.

Gabbay, D. M. & Guenthner, F. (eds.). (2001). *Handbook of Philosophical Logic 2nd Edition*, Vol. 3, Dordrecht: Kluwer Academic Publishers.

Gabbay, D. M., Kurucz, A., Wolter, F., Zakharyaschev, M. (2003). *Many-Dimensional Modal Logics: Theory and Applications*. Amsterdam: Elsevier.

Garson, J. W. (2006). *Modal Logic for Philosophers*. New York: Cambridge University Press.

Girle, R. (2000). *Modal Logics and Philosophy*. McGill-Queen's University Press.

Hilpinen, R. (ed.). (1971). *Deontic Logic: Introductory and Systematic Readings*. Dordrecht: D. Reidel Publishing Company.

Hilpinen, R. (ed.). (1981). *New Studies in Deontic Logic Norms, Actions, and the Foundation of Ethics*. Dordrecht: D. Reidel Publishing Company.

Kracht, M. (1999). *Tools and Techniques in Modal Logic*. Amsterdam: Elsevier.

Kracht, M. & Wolter, F. (1991). Properties of Independently Axiomatizable Bimodal Logics. *The Journal of Symbolic Logic*, vol. 56, no. 4, pp. 1469-1485.

Lewis, C. I. & Langford, C. H. (1932). *Symbolic Logic*. New York: Dover Publications. Second edition 1959.

Popkorn, S. (1994). *First Steps in Modal Logic*. Cambridge University Press.

Rescher, N. (ed.) (1967). *The Logic of Decision and Action*. Pittsburgh: University of Pittsburgh Press.

Segerberg, K. (1971). *An Essay in Classical Modal Logic*. 3 vols. Uppsala: University of Uppsala.

Zeman, J. J. (1973). *Modal Logic*: The Lewis-Modal Systems. Oxford: Clarendon Press.

Rönnedal, D. (2010). *An Introduction to Deontic Logic*. Charleston, SC.

Rönnedal, D. (2012). Bimodal Logic. *Polish Journal of Philosophy*. Vol. VI, No. 2, Fall 2012, pp. 71-93.

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

Åqvist, L. (1987). *Introduction to Deontic Logic and the Theory of Normative Systems*. Naples: Bibliopolis.

Åqvist L. (2002). Deontic Logic. In Gabbay & Guenthner (eds.) *Handbook of Philosophical Logic 2nd Edition*, Vol. 8, Dordrecht: Kluwer Academic Publishers, pp. 147-264.

Daniel Rönnedal
Department of Philosophy
Stockholm University
daniel.ronnedal@philosophy.su.se