Het Sleeping Beauty probleem als gedistribueerd systeem

Naam: Janneke de Bijl
Vak: Multiagent Systems
Datum: mei 2006

Inleiding
In dit paper behandel ik het Sleeping Beauty probleem. Ik zal eerst uiteenzetten welke vraag het probleem genereert, en welke twee (tegenstrijdige) antwoorden erop mogelijk zijn. Vervolgens zal ik de argumentatie weergeven die Elga gebruikt voor het verdedigen van een van beide antwoorden. Daarna bespreek ik een artikel van Halpern, die laat zien hoe het probleem gemodelleerd kan worden als gedistribueerd systeem. Ik zal ook zijn conclusies behandelen. Tot slot zal ik zelf een aantal opmerkingen maken over het probleem, en mijn eigenlijk conclusies trekken.

Het Sleeping Beauty probleem
Sleeping Beauty kan perfect rationeel redeneren en wordt twee dagen lang in slaap gehouden door onderzoekers. Tijdens die twee dagen wordt ze één of twee keer wakker gemaakt, afhankelijk van de uitkomst van een worp met een eerlijke munt. Bij kop wordt ze één keer wakker gemaakt en bij munt twee keer. Na het wakker worden wordt ze weer in slaap gebracht met een drug waardoor ze het wakker worden vergeet. Sleeping Beauty (SB) weet alles wat ik hierboven geschreven heb. Nu is de vraag: welke waarschijnlijkheid moet SB na de eerste ontwaking toekennen aan kop en aan munt?

Hierop lijken 2 redelijke antwoorden mogelijk te zijn. Het eerste antwoord is dat de kans op kop en de kans op munt allebei 1/2 is. Van tevoren ken je een kans van 1/2 toe, omdat je weet dat de munt eerlijk is. Door wakker te worden krijg je geen nieuwe informatie, want je wist van tevoren al dat je wakker gemaakt zou worden. De kans moet dus 1/2 blijven. Volgens het tweede antwoord moet je aan kop een kans van 1/3 toekennen, en aan munt een kans van 2/3. Bij herhaling van het experiment zal immers blijken dat 1/3 van de ontwakingen kop-ontwakingen zijn, en 2/3 van de ontwakingen munt-ontwakingen.

Elga
In '*Self-locating belief and the Sleeping Beauty problem*' verdedigt Elga de opvatting dat de kansen 1/3 en 2/3 zijn.
Als SB wakker wordt, zijn er 3 mogelijkheden:
H_1: Er is kop gegooid en het is maandag
T_1: Er is munt gegooid en het is maandag
T_2: Er is munt gegooid en het is dinsdag
Elga's argumentatie verloopt in 2 stappen. In de eerste stap toont hij aan dat de kans op T_1 gelijk moet zijn aan die op T_2, en in de tweede stap toont hij aan dat de kans op H_1 gelijk moet zijn aan die op T_1.

- Als je zou weten dat er munt is gegooid, dan zou je weten dat je je in T_1 of in T_2 bevindt. Omdat die twee toestanden voor SB ononderscheidbaar zijn, volgt uit het 'Principle of Indifference' dat SB aan beide toestanden dezelfde kansen zal toekennen. Dus gegeven dat SB weet dat er munt is gegooid, kent ze gelijke kansen toe aan T_1 en T_2. Omdat $P(T_1 - T_1 \vee T_2) = P(T_2 - T_1 \vee T_2)$ geldt volgens Elga ook $P(T_1) = P(T_2)$.
- Stel nu dat de munt pas wordt opgegooid na de eerste ontwaking op maandag. Dit moet niks uitmaken voor de kansen, want SB zou sowieso op maandag wakker gemaakt worden. Als je dan zou weten dat het maandag is, en je weet ook dat de munt daarna pas opgegooid wordt, dan is het duidelijk dat de kans op kop even groot is als de kans op munt. Omdat $P(H_1 - H_1 \vee T_1) = P(T_1 - H_1 \vee T_1)$ geldt volgens Elga ook $P(H_1) = P(T_1)$.

Uit deze twee resultaten volgt dat $P(H_1) = P(T_1) = P(T_2)$. Aangezien de som van de kansen 1 moet zijn is de kans op alledrie 1/3.

Elga geeft toe dat het wel erg ongebruikelijk is dat de kansen veranderen (vóór het experiment zijn ze immers 1/2) zonder dat SB nieuwe informatie krijgt. Hij verklaart dit doordat SB zich eerst in een situatie bevindt waarin haar eigen temporele locatie onbelangrijk is voor de kansen, en later in een situatie waarin haar eigen temporele locatie wel relevant is. Het verrassende is wel dat een dergelijke verandering plaatsvindt in een situatie waarin de agent volledig rationeel is en geen nieuwe informatie krijgt.

Voordat het experiment begint is het rationeel voor SB om te geloven dat de kans op kop, net als de kans op munt, 1/2 is. Maar van tevoren weet SB dat de kans na het wakker worden op maandag 1/3 zal zijn. Dit gaat in tegen het zogenaamde 'Reflection Principle' van Bas Van Fraassen. Dat principe houdt in dat als een rationele agent zeker weet dat hij morgen een bepaalde kans ergens aan zal toekennen, zonder dat zijn informatie dan veranderd zal zijn, hij diezelfde kans nu al moet toekennen. De Sleeping Beauty paradox vormt een nieuw soort tegenvoorbeeld tegen dit principe, aldus Elga.

Gedistribueerde Systemen

In het boek '*Epistemic Logic for AI and Computer Science*' wordt door Meyer en Van der Hoek beschreven hoe epistemische logica kan worden toegepast in gedistribueerde systemen. Hierbij maken zij gebruik van de theorie van Halpern en Moses. Een gedistribueerd systeem bestaat uit processoren die verbonden zijn door een communicatienetwerk. De lokale toestand van een processor is een functie van de begintoestand, ontvangen berichten en eventuele interne handelingen. De globale toestand van het systeem bestaat uit de lokale toestanden van de processoren. In het bijbehorende Kripke-model worden de relaties als volgt bepaald: Iedere processor weet in welke lokale toestand hij zelf is, maar niet in welke lokale toestanden de andere processoren zijn. Dus beschouwt die processor alle globale toestanden waarin hij dezelfde lokale toestand heeft als mogelijke werelden. In een bepaald punt geldt $K_i\phi$ desda ϕ waar is in alle werelden die processor i in dat punt voor mogelijk houdt.

In het boek worden ook mogelijke verfijningen van het model genoemd. Een daarvan houdt in dat toestanden vervangen worden door punten, die bestaan uit een run (een beschrijving van wat het systeem achtereenvolgens doet) en een tijdstip dat bepaald wordt door een globale klok. Een run dient om iets te kunnen zeggen over de verandering van kennis in het systeem, en is een opeenvolging van globale toestanden.

Deze verfijning van het model wordt verder uitgewerkt en gebruikt in het artikel '*Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems*', waarin Halpern het Sleeping Beauty probleem bespreekt. In het multiagent systems framework dat hij gebruikt zijn de processoren agents, en die agents bevinden zich op ieder tijdstip in een bepaalde lokale toestand. Die lokale toestand bevat alle informatie die de agent heeft. Wat hij hieraan toevoegt is dat het soms ook nuttig is om een 'omgeving' te modelleren, die alle relevante informatie bevat die niet in de lokale toestanden van de agents zit. In veel opzichten kan de omgeving ook als een agent worden beschouwd. Het hele systeem is in een bepaalde globale toestand, een tupel dat bestaat uit de toestand van de omgeving en de lokale toestanden van de agents. Om de veranderingen in het systeem weer te kunnen geven, maakt Halpern gebruik van de runs zoals ik die hierboven heb beschreven. Een run is een volledige beschrijving van een mogelijke manier waarop de globale toestand van het systeem

kan veranderen, ofwel een functie van tijd naar globale toestanden. Een paar van een run (r) en een tijdstip (m) wordt ook hier een punt genoemd. De punten die verbonden zijn door middel van equivalentierelaties worden ook wel informatieverzamelingen genoemd. Zo is K_i (r.m) de verzameling punten die door agent i voor mogelijk gehouden wordt op het punt (r,m). Dat zijn alle punten waarin i dezelfde lokale toestand heeft, want die punten zijn voor i ononderscheidbaar van de werkelijke toestand. Deze *K* dient niet verward te worden met de K van kennis. $K_i\phi$ is een bewering over de kennis van een bepaalde agent, terwijl K_i (r.m) een verzameling punten aanduidt die voor een agent ononderscheidbaar zijn.

Halpern stelt dat dit multiagent systems framework zeer geschikt is om een aantal belangrijke aannames te modelleren, zoals perfect recall en synchroniteit. Een systeem is synchroon als voor alle agents geldt dat ze op ieder tijdstip m weten dat het tijdstip m is, dus alle punten die ze voor mogelijk houden op tijdstip m zijn punten waarop het tijdstip ook m is. Een agent heeft perfect recall als zijn lokale toestand altijd groeit wanneer er nieuwe informatie wordt toegevoegd. Dit betekent dat een agent vanuit zijn huidige lokale toestand zijn volledige geschiedenis van lokale toestanden kan reconstrueren. In het model houdt dit in dat de agent in de toestanden die voor hem ononderscheidbaar zijn dezelfde opeenvolging van lokale toestanden heeft. Intuïtief wil dit zeggen dat de agent niks vergeet.

Zoals we zullen zien gebruikt Halpern dit framework om de opvatting te verdedigen dat het SB probleem ontstaat doordat de tijd (voor SB) niet synchroon is, en niet zoals sommigen beweren doordat er sprake is van ‘imperfect recall’. Het is dus niet het feit dat SB op dinsdag vergeten is dat ze eerder al wakker was dat de paradox veroorzaakt, maar het feit dat ze niet weet welke dag het is. Laten we nu eerst kijken naar hoe Halpern het SB probleem modelleert.

Sleeping Beauty als gedistribueerd systeem

De belangrijkste agent in het SB probleem is uiteraard SB zelf. In haar lokale toestand vóór het experiment zit onder andere de informatie over hoe het experiment werkt, en bijvoorbeeld dat het zondag is. Bij het wakker worden bevat haar lokale toestand nog steeds informatie over hoe het experiment werkt, maar weet ze niet of het maandag of dinsdag is (maar wel dat het maandag of dinsdag is). De informatie die SB niet heeft maar ‘wij’ wel, kan bevat worden door de omgeving. Dat is hier de informatie over welke dag het is en of er kop of munt is gegooid. De omgeving kan hier ook heel goed beschouwd worden als de lokale toestand van de onderzoekers die het experiment uitvoeren.

Volgens Halpern zijn er verschillende manieren waarop het SB probleem als gedistribueerd systeem gemodelleerd kan worden. Maar aangezien de onderzoeker die het experiment uitvoert verder geen rol speelt, kan het het beste als een single-agent probleem worden beschouwd. We kunnen de volgende toestanden onderscheiden: de toestand vóór het experiment, de toestand waarin SB wakker gemaakt is, en de toestand na het experiment. De toestanden waarin SB slaapt zouden ook gemodelleerd kunnen worden, maar die zijn volgens Halpern niet relevant voor het probleem, dus laat hij ze weg. Dit leidt tot het model R_1 met twee runs, run 1 voor kop en run 2 voor munt. Dit ziet er in Halperns artikel als volgt uit:

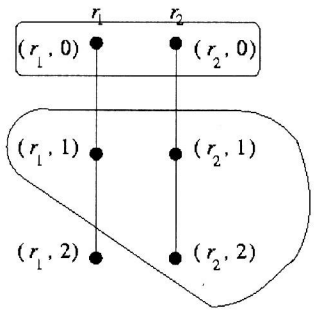

Figure 1: The Sleeping Beauty problem, captured using $\mathcal{R}_1$.

Zoals uit het plaatje blijkt zijn alleen de eerste 3 punten weergegeven (bij run 2 is er eigenlijk nog een vierde punt). De punten die relevant zijn voor het probleem, zijn de toestanden waarin SB wakker gemaakt is. Die drie punten zijn in het plaatje omlijnd, en vormen een informatieverzameling, omdat ze voor SB ononderscheidbaar zijn.

Nu we een model hebben, kunnen we kijken of er in het model sprake is van perfect recall en/of synchroniteit. We zien dat model R_1 niet synchroon is, want SB houdt op maandag ook punten voor mogelijk waarop het dinsdag is. Wel heeft SB volgens Halpern perfect recall in dit model. Dit komt doordat hij de toestanden waarin SB slaapt niet heeft gemodelleerd. Toch blijkt het SB probleem duidelijk uit het model. Dat betekent dat het probleem ontstaat door asynchroniteit, en niet door imperfect recall.

Kansen toekennen

In R_1 is de kansverdeling over de runs duidelijk: de kans op kop is evenals de kans op munt 1/2. Het is immers een eerlijke munt. Het probleem ontstaat zodra we kansen willen toekennen aan $(r_1,1)$, $(r_2,1)$ en $(r_2,2)$. De beginkans van de twee runs is niet voldoende om te weten hoe die verdeeld moeten worden over die drie punten. Het spreekt voor zich dat de kansen van de punten gerelateerd moet zijn aan de kansen van de runs, maar hoe? Bij synchrone gevallen is het verdelen van de kansen van de runs over punten erg eenvoudig, omdat de kansen van de punten dan gelijk zijn aan de kansen van de runs. In asynchrone gevallen zijn de punten in K_i (r.m) niet allemaal m-punten, waardoor niet duidelijk is hoe je de kansen moet verdelen. Volgens Halpern zijn er twee redelijke benaderingen mogelijk, die overeenkomen met de twee antwoorden die Elga bespreekt. De eerste benadering noemt Halpern de HT-benadering, omdat hij een methode toepast die hij samen met M. Tuttle ontwikkelde. De tweede benadering noemt hij de Elga-benadering, omdat die uitkomt op het antwoord dat Elga verdedigt.

- HT-benadering:

  Volgens deze benadering wordt de kans van de run geprojecteerd op de punten op de run. In dit geval geldt dat beide runs kans 1/2 hebben, en de 1/2 van run 1 wordt dus volledig geprojecteerd op punt $(r_1, 1)$. De 1/2 van r_2 wordt geprojecteerd op de twee punten op die op run 2 liggen, maar het is niet duidelijk hoe die daarover verdeeld moet worden. De HT-benadering zou zeggen dat het niet mogelijk is om daar verder uitspraken over te doen, dus het enige dat men weet is dat die kansen samen 1/2 zijn. Een alternatief hiervoor is om het Principle of Indifference toe te passen, en te zeggen dat beide punten kans 1/4 krijgen.
- Elga-benadering:

  Deze benadering gaat uit van de verhoudingen van kansen tussen punten. Voor ieder paar punten in verschillende runs moet de relatieve waarschijnlijkheid

hetzelfde zijn als die van de globale toestanden. In dit geval is de waarschijnlijkheid van beide runs 1/2, dus moet de verhouding tussen (r_1,1) en (r_2,1) 1:1 zijn, evenals de verhouding tussen (r_1,1) en (r_2,2). Dit zou betekenen dat de kansen alledrie even groot zijn, en dus 1/3.

Dus terwijl de HT-methode de kansen van een run verdeelt over de punten in de informatieverzameling die op die run liggen, kent de Elga-methode aan ieder punt op de informatieverzameling relatief dezelfde kans toekent als aan de run. Uiteraard worden de resultaten bij beide methoden genormaliseerd zodat de som van de kansen 1 is.

HT-benadering of Elga-benadering?

De vraag is nu welke van deze twee benaderingen het beste is. Volgens Halpern maakt Elga een belangrijke fout in zijn redenering. Het gaat om de stap waarbij Elga uit het feit dat als je zou ontdekken dat het maandag is, je evenveel kans aan kop en munt zou toekennen, concludeert dat de kans op H1 even groot moet zijn als die op T1. Volgens Halpern mag je de kans op kop gegeven dat je leert dat het maandag is niet identificeren met de kans op kop gegeven dat het maandag is. Deze opvatting ondersteunt hij door het volgende voorbeeld: Als een man denkt dat zijn vrouw slimmer is dan hij, dan acht hij de kans klein dat hij erachter komt als ze vreemdgaat. Dus de kans op Y (dat hij erachter komt dat ze vreemdgaat) gegeven X (dat ze vreemdgaat), is heel klein. Maar de kans op Y gegeven dat hij leert dat X is 1. Dat Elga om de verkeerde reden concludeert dat het antwoord 1/3 – 2/3 moet zijn, betekent echter nog niet dat zijn antwoord onjuist is, aldus Halpern.

In het artikel komt Halpern niet met een oplossing voor de paradox. Hij legt uit dat ook de frequentie-interpretatie, waarbij waarschijnlijkheden worden beschouwd als relatieve frequenties, geen uitkomst biedt in het SB probleem. Normaal gesproken kun je kijken naar het aantal keren dat een experiment wordt uitgevoerd om de kans te bepalen, maar het probleem is dat hier niet duidelijk is wat als een keer beschouwd moet worden. Wanneer men het hele experiment als een keer beschouwd, zal het antwoord 1/2 –1/2 zijn, en wanneer men iedere ontwaking als een keer beschouwd, zal er 1/3 - 2/3 uitkomen.

Halpern maakt nog wel een aantal belangrijke punten in zijn artikel. Hij beweert dat het Reflection Principle van Van Fraassen gebaseerd is op een aantal aannames. Een van die aannames is dat er een vaste verzameling mogelijke werelden zou zijn. Bij het SB-probleem is dat niet het geval. Verder gaat het Reflection Principle wel op in situaties met perfect recall en synchroniteit, maar is het nog maar de vraag of het bij afwezigheid daarvan geldt. Hij argumenteert uiteindelijk dat het niet geldig is in situaties met imperfect recall, maar dat een aangepaste versie van het principe volgens de HT-benadering wel kan worden toegepast op asynchrone systemen. Volgens de Elga-methode is dit niet mogelijk. Halpern concludeert verder dat onze intuities op het gebied van waarschijnlijkheid ons in de steek laten op het moment dat het gaat om situaties met imperfect recall of asynchroniteit.

Evaluatie

Volgens mij is het heel zinvol om het SB probleem te modelleren als een gedistribueerd systeem. De begintoestand van SB bevat de informatie over hoe het experiment werkt, en het feit dat het zondag is. Haar lokale toestand bij het ontwaken bevat nog steeds de informatie over het experiment, en verder het feit dat het ‘maandag of dinsdag’ is. Deze lokale toestand is hetzelfde bij alledrie de ontwakingen. Omdat de momenten waarop SB slaapt niet gemodelleerd worden, is er,

zoals Halpern beweert, sprake van perfect recall. Dit lijkt vreemd, omdat SB zich op dinsdag niet meer herinnert dat ze maandag wakker was. Maar omdat de toestanden waarin ze slaapt niet gemodelleerd worden, bevindt ze zich op die dinsdag in dezelfde lokale toestand als op die maandag, en kan ze wel degelijk haar geschiedenis van lokale toestanden reconstrueren. De enige lokale toestand die eraan vooraf ging is dan immers de lokale toestand die ze vóór het experiment had, en die herinnert ze zich nog. Wanneer het slapen ook gemodelleerd wordt kan SB wanneer ze ontwaakt haar geschiedenis van lokale toestanden niet reconstrueren; ze weet immers niet of ze één of twee keer heeft geslapen. In het model van Halpern is wel sprake van perfect recall en niet van synchroniteit, terwijl het SB probleem er nog steeds duidelijk uit naar voren komt. Hiermee heeft Halpern dus aangetoond dat het SB probleem ontstaat door asynchroniteit. Ik denk dat dit zeker kan helpen bij het zoeken naar een oplossing voor het probleem. Nu kan het onderzoek gericht worden op het toekennen van kansen bij asynchrone systemen, in plaats van bij systemen met imperfect recall.

Een minder punt uit Halperns artikel is zijn kritiek op Elga's redenering. Het voorbeeld dat hij gebruikt om te laten zien dat de kans op X gegeven Y niet gelijk is aan de kans op X gegeven dat je leert dat Y, vind ik niet sterk. Halpern gebruikt de volgende analogie:

- De kans op kop, gegeven dat het maandag is, is niet gelijk aan de kans op kop gegeven dat je leert dat het maandag is want
- de kans dat een man erachter komt dat zijn vrouw vreemdgaat, gegeven dat zijn vrouw vreemdgaat, is niet gelijk aan de kans dat hij erachter komt dat zijn vrouw vreemdgaat, gegeven dat hij leert dat zijn vrouw vreemdgaat.

In het voorbeeld over de ontrouwe echtgenote is X 'de man komt erachter dat zijn vrouw vreemdgaat' en Y 'zijn vrouw gaat vreemd'. Maar hieruit blijkt al dat X herschreven kan worden tot 'de man komt erachter dat Y'. Als we dit nu gebruiken in het herschrijven van de redenering wordt dit:

- De kans dat iemand erachter komt dat X gegeven dat X is niet gelijk aan de kans dat iemand erachter komt dat X gegeven dat hij leert dat X.

Aangezien de kans op dat laatste uiteraard 1 is, kan dit weer geherformuleerd worden tot:

- Als iets zo is dan is de kans dat iemand daar ook achter komt niet 1.

ofwel:

- Niet iedereen komt altijd overal achter.

Dit is uiterst triviaal, en het lijkt me duidelijk dat Elga's redenering wel meer inhoudt dan dit. De redenering van Elga gaat als volgt:

- De kans op H1 is even groot als de kans op T1 want
- de kans op H1 gegeven dat je leert dat het maandag is, is even groot als de kans op T1 gegeven dat je leert dat het maandag is.

Omdat het leren in beide delen van de tweede zin zit, kan dat in de 'X' gestopt worden. De redenering heeft dan de volgende structuur:

- De kans op Y is even groot als de kans op Z want
- de kans op Y gegeven X is even groot als de kans op Z gegeven X.

Hierbij is X dus 'je leert dat het maandag is'. Het tegenvoorbeeld van Halpern gaat dus duidelijk niet op, omdat die redenering een hele andere structuur heeft dan Elga's redenering.

Verder wil ik opmerken dat de vraag die leidt tot het SB probleem niet optimaal geformuleerd is. De vraag luidt: welke kansen moet je toekennen na de

eerste keer wakker worden? In deze vraag gaat het om de kansen die SB zal toekennen na de eerste keer wakker worden, en dat zal ze natuurlijk doen vanuit haar lokale toestand op dat moment. En in die lokale toestand zit helemaal niet besloten welke dag het is. Met die vraag wordt dus bedoeld: welke kansen zal SB toekennen na de eerste keer ontwaken, terwijl ze niet weet dat het de eerste keer is. In die vraag worden twee lokale toestanden met elkaar vermengd: die van SB, die bij het ontwaken gebrekkige informatie heeft, en die van de onderzoeker, die wel weet welke dag het is en wat de uitkomst van de muntworp is. Het vermelden van het feit dat het gaat om de eerste ontwaking in de vraag lijkt me even irrelevant als het in de vraag vermelden dat er bijvoorbeeld kop is gegooid: Stel dat er kop is gegooid, welke kansen zal SB dan bij de eerste ontwaking toekennen? Uiteraard dezelfde kansen als wanneer er munt is gegooid, want SB heeft die informatie dan helemaal niet. Om dezelfde reden zal ze na de eerste ontwaking altijd dezelfde kansen toekennen als na de tweede ontwaking. Het feit dat het de eerste keer is kan dus net zo goed weggelaten worden uit de vraag. Laten we de vraag dus herformuleren tot: welke kansen zal SB toekennen na het ontwaken?

Het lijkt erop dat we te weinig informatie over het experiment hebben om deze vraag te kunnen beantwoorden. Een mogelijke interpretatie van de informatie die we hebben, is dat die vraag haar iedere keer bij het ontwaken zal worden gesteld. In dat geval geldt het argument dat 1/3 van de ontwakingen kop-ontwakingen zijn, en 2/3 van de ontwakingen munt-ontwakingen. Dus zal SB bij iedere ontwaking die kansen toekennen. Dit lijkt in te gaan tegen het Reflection Principle van Van Fraassen. Van tevoren weet SB al dat ze op maandag wakker gemaakt zal worden, maar na het ontwaken veranderen de kansen, schijnbaar zonder dat SB nieuwe informatie krijgt. Volgens mij klopt dit niet, omdat de informatie van SB wel degelijk verandert. Vóór het experiment weet SB wel dat ze op maandag wakker gemaakt zal worden, maar wanneer ze op maandag wakker is weet ze helemaal niet dat het maandag is. Op zondag weet ze dat de kansen 1/2 zijn, en weet ze ook dat ze op maandag wakker gemaakt zal worden. Toch heeft ze bij het ontwaken op maandag ‘niewe’ informatie, namelijk de ‘informatie’ dat het misschien dinsdag is. Dit is eigenlijk een vermindering van informatie; vooraf weet ze dat het dan maandag zal zijn, maar bij het ontwaken weet ze alleen nog dat het maandag of dinsdag is. Het klopt dat SB van tevoren al weet dat ze minstens een keer wakker wordt gemaakt, maar wanneer ze de eerste keer wakker wordt ‘weet’ ze dat het misschien de tweede keer is. Het lijkt erop dat het woord ‘eerste’ in de oorspronkelijke formulering van de vraag die het SB probleem genereert bijdraagt aan de misvatting dat dit antwoord indruist tegen het Reflection Principle.

Een andere mogelijke interpretatie is dat SB maar een keer per experiment wordt gevraagd om kansen toe te kennen. Bij kop zou dat dan altijd op maandag gebeuren, en bij munt op maandag of op dinsdag. Maar aangezien de vraag oorspronkelijk ging over de eerste ontwaking, lijkt het zo te zijn dat de vraag in ieder geval gesteld wordt na de eerste ontwaking. Dat zou betekenen dat de vraag volgens deze interpretatie, ook als er munt is gegooid, alleen op maandag zal worden gesteld. Als de vraag maar één keer per experiment wordt gesteld, dan is de kans op kop uiteraard even groot als de kans op munt. Het verschil tussen deze twee interpretaties komt overeen met Halperns opmerking over de frequentie-interpretatie van waarschijnlijkheid. Deze kan inderdaad niet worden gebruikt om het SB probleem op te lossen, omdat niet duidelijk is wat als een ‘keer’ dient te worden beschouwd.

Zolang het experiment niet duidelijk is omschreven, kan SB niet weten welke van deze twee interpretaties het meest waarschijnlijk is. Het lijkt me dan ook het

meest rationeel voor SB om het Principle of Indifference toe te passen, en er vanuit te gaan dat beide varianten van het experiment even waarschijnlijk zijn. Dit betekent dat de kans die ze zal toekennen aan kop 1/2 + 1/3 gedeeld door 2, en de kans op munt 1/2 + 2/3 gedeeld door 2 zal zijn. De kans die ze aan kop toekent is dan 5/12 en de kans die ze aan munt toekent is 7/12.

Conclusies

Het is zinvol om het Sleeping Beauty probleem te modelleren als gedistribueerd systeem. Halpern toont met behulp van een model aan dat het probleem niet ontstaat door imperfect recall, maar door asynchroniteit. In asynchrone systemen is het lastig om kansen toe te kennen, omdat niet duidelijk is hoe de kansen van de runs over de punten verdeeld dienen te worden. De vraag die het SB probleem genereert is niet duidelijk geformuleerd, doordat informatie uit twee verschillende lokale toestanden gecombineerd wordt. Het SB probleem ontstaat doordat er te weinig informatie is over het experiment. Beide antwoorden die Elga in zijn tekst bespreekt zijn dan ook verdedigbaar. Toch is het probleem minder groot dan Elga en Halpern denken, doordat geen van beide antwoorden indruisen tegen het Reflection Principle van Van Fraassen.

OVERZICHT VAN DE GEBRUIKTE LITERATUUR

- Meyer, J.-J. Ch. en Van der Hoek, W., *Epistemic Logic for AI and Computer Science*, Cambridge Univerity Press, 1995
- Elga, A., *Self-locating belief and the Sleeping Beauty problem*, Analysis, 60 (2), 143-147, 2000
- Halpern, J.Y., *Sleeping Beauty Reconsidered: Conditioning and Reflection in Asynchronous Systems*, Ninth International Conference on Principles of Knowledge Representation and Reasoning (KR 2004), 12-22, 2004