

## Stalnaker's Argument

(This is a supplement to 'Countable Additivity, Dutch Books and the Sleeping Beauty Problem')

Stalnaker (2008a) suggests an argument that can be stated thus:

Let  $t$  be the time at which Beauty awakens on Monday morning. Upon awakening, Beauty learns something that is relevant to Heads: while she doesn't know when  $t$  is, she does learn, of  $t$ , that she is awake *then*. Having learned this, she knows that one of the following three possibilities obtains:

T1\*: Tails and  $t$  is on Monday.

T2\*: Tails and  $t$  is on Tuesday.

H1\*: Heads and  $t$  is on Monday.

But if Beauty is to be represented as learning, on Monday morning, that she is awake at  $t$ , then she must be represented as not yet knowing this on Sunday evening, and hence as having not yet ruled out the following possibility:

H2\*: Heads and  $t$  is on Tuesday.

On Sunday evening, Beauty should regard each of the four above possibilities as equally probable. And on Monday morning, she should conditionalize on the negation of the fourth possibility, with the result that her credence in H1\*, and hence in Heads, is one-third.

Unfortunately, while Stalnaker says that "all parties to the debate should be able to agree" about how Beauty's "degrees of belief should be apportioned on Sunday" among these four possibilities, he doesn't indicate how this consensus should be reached. However, elsewhere (2008b) he offers a clue about the kind of principle he wants to invoke:

[Beauty] is in a position that is like [that of] a person who is asked to make a probability judgment about the following situation: Two people are walking together in Central Park. Call them  $A$  and  $B$ . How probable do you think it is that  $B$  is wearing running shoes, given that  $A$  is? ...[H]owever she answers the question, if she is rational, she had better answer this follow-up question in the same way: Now consider the conditional degree of belief that it is  $A$  who is wearing running shoes, on the condition that  $B$  is. The reason the questions have to be answered the same way is that they are really the same question, since ' $A$ ' and ' $B$ ' are just the theorist's labels for two generically identified people ... The only indifference principle that I want to rely on is the principle used in this example, requiring that the two questions about conditional degree of belief be answered in the same way for generically described situations. In the SB case, this principle applies on Sunday to the possibilities that will be distinguished on Monday and Tuesday, but that can be described only generically on Sunday.

On some ways of formalizing the indifference principle Stalnaker alludes to, this principle would, by itself, conflict with CA. The interesting question, however, is whether CA conflicts with a restricted,

finitistic version of this indifference principle, together with the other assumptions needed to derive the one-third solution. We can formulate a restricted version of the indifference principle as follows:

**Arbitrary Time Indifference.** Suppose  $T$  is a finite set consisting of  $n$  times,  $t_1, t_2, \dots, t_n$ , such that Beauty represents the elements of  $T$  simply as arbitrary times. And for some proposition,  $p$ , and some predicate,  $A$ , suppose that, conditional on  $p$ , the set of propositions  $\{A(t_1), A(t_2), \dots, A(t_n)\}$  constitutes a partition. In this case, Beauty should have equal credence, conditional on  $p$ , in each of the propositions in  $T$ .

We can derive the one-third solution from the above principle, together with Finite Additivity, the Restricted Principal Principle, and the following three assumptions:

**Prior Admissibility.** In any Sleeping Beauty problem, on the evening prior to the first experimental awakening, Beauty knows that she has no information that is inadmissible in relation to any of the hypotheses in  $S$ .

**De Re Conditionalization.** In any Sleeping Beauty problem, if  $t$  is the time at which Beauty first awakens, then her credences at  $t$  should be the same as they would be if she had initially represented  $t$  as an arbitrary time, and then conditionalized on the proposition that she undergoes an experimental awakening at  $t$ .

**Arbitrary Time Independence.** In any Sleeping Beauty problem, for any hypothesis  $h$  in  $S$ , where proposition  $p$  is a specification of the locations of times that Beauty represents simply as arbitrary times, Beauty's credence in  $h$  given  $p$  should be equal to her unconditional credence in  $h$ .

In order to derive the one-third solution, let  $t$  represent the time when Beauty awakens on Monday morning, and let  $t'$  represent any other time which Beauty represents, on both on Sunday evening and on Monday morning, simply as an arbitrary time. Let  $P$  be Beauty's credence function when she awakens on Monday morning, and let  $P'$  be the credence function she would have on Sunday evening, assuming that at that time she also represented  $t$  simply as an arbitrary time. Let the propositions  $H1'$  and  $X$  be defined as follows:

$H1'$ . Heads and  $t'$  is on Monday.

$X$ .  $t$  and  $t'$  are the two times when Beauty would undergo experimental awakenings given Tails.

And let  $Y$  be the following proposition, which is entailed by  $X$ :

$Y$ .  $t'$  is one of the times when Beauty would undergo experimental awakenings given Tails, and  $t'$  does not coincide with  $t$ .

Conditional on (Heads & X), the set {H1\*, H1'} constitutes a partition. And since the propositions in this set differ only with respect to times which, relative to  $P^-$ , are represented simply as arbitrary times, it follows from Arbitrary Time Indifference that

$$P^-(H1^*|Heads \& X) = P^-(H1'|Heads \& X)$$

And conditional on (Heads & X), H1' is equivalent to H2\*. Hence,

$$P^-(H1^*|Heads \& X) = P^-(H2^*|Heads \& X)$$

Since, conditional on X, Heads is equivalent to the disjunction of H1\* and H2\*, we may infer

$$P^-(H1^* | X) = P^-(H2^* | X) = \frac{1}{2} P^-(Heads | X)$$

But since X is a specification of the locations of times which, relative to  $P^-$ , are represented simply as arbitrary times, it follows from Arbitrary Time Independence that

$$P^-(Heads | X) = P^-(Heads)$$

And from the Principal Principle, together with Prior Admissibility

$$P^-(Heads) = Ch(Heads) = 1/2$$

Hence, from the last three lines,

$$P^-(H1^* | X) = P^-(H2^* | X) = \frac{1}{4} \tag{\alpha}$$

We can now determine Beauty's Monday morning credence in Heads. Since Y is a specification of the locations of a time which, on Monday morning, Beauty represents simply as an arbitrary time, it follows from Arbitrary Time Indifference that

$$P(Heads) = P(Heads|Y)$$

And where W is the proposition that Beauty undergoes an experimental awakening at t, it follows from De Re Conditionalization that

$$P(Heads|Y) = P^-(Heads|W \& Y)$$

And since (W & Y) entails X,

$$P(\text{Heads} | W \& Y) = P(\text{Heads} | W \& X \& Y) = \frac{P^-(\text{Heads} \& W | X \& Y)}{P^-(W | X \& Y)}$$

But conditional on  $(X \& Y)$ ,  $(\text{Heads} \& W)$  is equivalent to  $H1^*$ , and  $W$  is equivalent to the negation of  $H2^*$ . Hence, from  $(\alpha)$ ,

$$\frac{P^-(\text{Heads} \& W | X \& Y)}{P^-(W | X \& Y)} = \frac{P^-(H1^* | X)}{P^-(\neg H2^* | X)} = \frac{1/4}{3/4} = \frac{1}{3}$$

And so it follows, from the last four lines, that  $P(\text{Heads}) = 1/3$ .

I will now argue that the premises of the above argument conflict with CA. In any Sleeping Beauty problem defined by a partition,  $S$ , let  $i$  and  $j$  be any two hypotheses in  $S$ . If  $N(j) \geq N(i)$ , then let 'A' denote  $i$ , and let 'B' denote  $j$ ; otherwise, let 'A' denote  $j$  and let 'B' denote  $i$ . Thus,  $N(B) \geq N(A)$ . Let  $t_0$  be the time of Beauty's first experimental awakening. And let  $t_2, t_3, \dots, t_{N(B)}$  be any other times such that, both upon first awakening and on the previous evening, Beauty represents these times simply as arbitrary times. Let  $P$  be Beauty's credence function when she first awakens, and let  $P^-$  be the credence function she would have the previous evening, assuming that at that time she also represented  $t_1$  simply as an arbitrary time. Let propositions  $X'$  and  $Y'$  be defined as follows:

- $X'$ . The times  $t_1$  through  $t_{N(B)}$  are the  $N(B)$  times when Beauty would undergo an experimental awakening given  $B$ .
- $Y'$ . The times  $t_2$  through  $t_{N(B)}$  are all distinct times at which Beauty would undergo an experimental awakening given  $B$ , and none of these times coincides with  $t_1$ .

And for any two integers,  $m$  and  $n$ , between 1 and  $N(B)$ , let us define the proposition  $a_m^n$  as follows:

$a_m^n$ . Hypothesis  $A$  is true, and  $t_m$  is the time of the  $n$ th experimental awakening.

Now consider the following matrix:

$$\begin{array}{cccc} a_1^1 & a_1^2 & \dots & a_1^{N(B)} \\ a_2^1 & a_2^2 & \dots & a_2^{N(B)} \\ \dots & \dots & \dots & \dots \\ a_{N(B)}^1 & a_{N(B)}^2 & \dots & a_{N(B)}^{N(B)} \end{array}$$

We can now determine the probabilities of the propositions in this row, relative to  $P^-$ . Conditional on  $(A \& X')$ , every row, and every column, of this matrix constitutes a partition. And the propositions in any given column differ from one another only with respect to times which, relative to  $P^-$ , are represented

simply as arbitrary times. And so it follows from Arbitrary Time Indifference that, conditional on  $(A \& X')$ , each of the propositions in any given column must have the same probability, and these probabilities must sum to one. Therefore, every proposition in any given row must likewise have the same probability, conditional on  $(A \& X')$ .

Consider the first row. As we have seen, conditional on  $(A \& X')$ , all the propositions in this row have the same probability. And, conditional on  $X'$ , the disjunction of these propositions is equivalent to  $A$ . Hence, conditional on  $X'$ , Beauty's credences in these propositions must be equal, and must sum to her credence in  $A$ . Thus, since there are  $N(B)$  propositions in this row,

$$\text{For all } n \text{ s.t. } 1 \leq n \leq N(B), \quad P^-(a_1^n | X') = \frac{P^-(A | X')}{N(B)}$$

Applying Arbitrary Time Independence, the Principal Principle, and Prior Admissibility, we obtain

$$\text{For all } n \text{ s.t. } 1 \leq n \leq N(j), \quad P^-(a_1^n | X') = \frac{P^-(A)}{N(B)} = \frac{Ch(A)}{N(B)} \quad (\alpha')$$

We can now determine what Beauty's credences should be when she first awakens. Since  $Y'$  is a specification of the locations of times which, upon first awakening, Beauty represents simply as arbitrary times, it follows from Arbitrary Time Independence that  $P(A) = P(A | Y')$ . Hence, where  $W'$  is the proposition that Beauty undergoes an experimental awakening at  $t_i$ , it follows from De Re Conditionalization that  $P(A) = P^-(A | W' \& Y')$ . By similar reasoning,  $P(B) = P^-(B | W' \& Y')$ . Hence,

$$P(A)P^-(B | W' \& Y') = P(B)P^-(A | W' \& Y')$$

And since  $(A \& W' \& Y')$  and  $(B \& W' \& Y')$  each entails  $X'$ ,

$$P(A)P^-(B \& W' | X' \& Y') = P(B)P^-(A \& W' | X' \& Y')$$

But conditional on  $(X' \& Y')$ ,  $(B \& W')$  is equivalent to  $B$ , and  $(A \& W')$  is equivalent to the disjunction  $a_1^1 \vee a_1^2 \vee \dots \vee a_1^{N(a)}$ . Hence,

$$P(A)P^-(B | X') = P(B)P^-(a_1^1 \vee a_1^2 \vee \dots \vee a_1^{N(a)} | X')$$

And so, from  $(\alpha')$ , by Finite Additivity,

$$P(A)P^-(B | X') = P(B)N(A)Ch(A) / N(B)$$

And by Arbitrary Time Independence, the Restricted Principal Principle, and Prior Admissibility,

$$P^-(B|X') = P^-(B|X') = Ch(B)$$

Hence, from the last two lines

$$P(A)N(B)Ch(B) = P(B)N(A)Ch(A) \quad \text{whenever } N(i), N(i) \in Z$$

Now recall that  $A$  and  $B$  are simply the arbitrary hypotheses,  $i$  and  $j$ , belonging to  $S$ . Because of the symmetry of the above equation, we can infer that, regardless of whether  $A = i$  and  $B = j$ , or vice versa,

$$P(i)N(j)Ch(j) = P(j)N(i)Ch(i) \quad \text{whenever } N(i), N(i) \in Z$$

Thus, the premises of the De Re Conditionalization argument entail the GTP.

## References

Stalnaker, R. (2008a). *Our Knowledge of the Internal World*. Oxford: Oxford University Press.

Stalnaker, R. (2008b). Comment on “Stalnaker on Sleeping Beauty” *Brain Pains: a group blog in the philosophy of mind*. <http://brainpains.com/wordpress/?p=54> (accessed Oct. 12, 2009).