SOME PROBLEMS FOR CONDITIONALIZATION AND REFLECTION*

I will present five puzzles that show that rational people can update their degrees of belief in manners that violate Bayesian conditionalization and Bas van Fraassen's reflection principle. I will then argue that these violations of conditionalization and reflection are due to the fact that there are two as yet unrecognized ways in which the degrees of belief of rational people can develop.

I. TWO ROADS TO SHANGRI LA

Every now and then, the guardians to Shangri La will allow a mere mortal to enter that hallowed ground. You have been chosen because you are a fan of the Los Angeles Clippers. But there is an ancient law about entry into Shangri La: you are only allowed to enter, if, once you have entered, you no longer know by what path you entered. Together with the guardians, you have devised a plan that satisfies this law. There are two paths to Shangri La, the Path by the Mountains, and the Path by the Sea. A fair coin will be tossed by the guardians to determine which path you will take: if heads you go by the Mountains, if tails you go by the Sea. If you go by the Mountains, nothing strange will happen: while traveling you will see the glorious Mountains, and even after you enter Shangri La, you will forever retain your memories of that Magnificent Journey. If you go by the Sea, you will revel in the Beauty of the Misty Ocean. But, just as you enter Shangri La, your memory of this Beauteous Journey will be erased and be replaced by a memory of the Journey by the Mountains.

Suppose that in fact you travel by the Mountains. How will your degrees of belief develop? Before you set out your degree of belief in heads will be $\frac{1}{2}$. Then, as you travel along the Mountains and you gaze upon them, your degree of belief in heads will be one. But then, once you have arrived, you will revert to having degree of belief $\frac{1}{2}$ in heads. For you will know that you would have had the memories that you have either way, and hence you know that the only relevant information that you have is that the coin was fair.

This seems a bizarre development of degrees of belief. For as you

*I would like to thank John Collins, Adam Elga, John Hawthorne, Isaac Levi, Barry Loewer, and Tim Maudlin for extensive and crucial comments and discussions on earlier versions of this article.

0022-362X/03/0007/356–70

are traveling along the Mountains, you know that your degree of belief in heads is going to go down from one to ½. You do not have the least inclination to trust those future degrees of belief. Those future degrees of belief will not arise because you will acquire any evidence, at least not in any straightforward sense of "acquiring evidence." Nonetheless, you think you will behave in a fully rational manner when you acquire those future degrees of belief. Moreover, you know that the development of your memories will be completely normal. It is only because something strange would have happened to your memories had the coin landed tails that you are compelled to change your degree of belief to ½ when that counterfactual possibility would have occurred.

II. THE PRISONER

You have just been returned to your cell on death row, after your last supper. You are to be executed tomorrow. You have made a last minute appeal to President George W. Bush for clemency. Since Dick Cheney is in the hospital and cannot be consulted, George W. will decide by flipping a coin: heads you die, tails you live. His decision will be made known to the prison staff before midnight. You are friends with the prison officer who will take over the guard of your cell at midnight. He is not allowed to talk to you, but he will tell you of Bush's decision by switching the light in your cell off at the stroke of midnight if it was heads. He will leave it on if it was tails. Unfortunately you do not have a clock or a watch. All you know is that it is now 6 PM since that is when prisoners are returned to their cells after supper. You start to reminisce and think fondly of your previous career as a Bayesian. You suddenly get excited when you notice that there is going to be something funny about the development of your degrees of belief. Like anybody else, you do not have a perfect internal clock. At the moment you are certain that it is 6 PM, but as time passes your degrees of belief are going to be spread out over a range of times. What rules should such developments satisfy?

Let us start on this problem by focusing on one particularly puzzling feature of such developments. When in fact it is just before midnight, say 11:59 PM, you are going to have a certain, nonzero, degree of belief that it is now later than midnight. Of course, at 11:59 PM the light in your cell is still going to be on. Given that at this time you will have a nonzero degree of belief that it is after midnight, and given that in fact you will see that the light is still on, you will presumably take it that the light provides some evidence that the outcome was tails. Indeed, it seems clear that as it gets closer to midnight, you will monotonically increase your degree of belief in tails. Moreover you

know in advance that this will happen. This seems puzzling. Of course, after midnight, your degree of belief in tails will either keep on increasing, or it will flip to zero at midnight and stay there after midnight. But that does not diminish the puzzlement about the predictable and inevitable increase in your degree of belief in tails prior to midnight. In fact, it seems that this increase is not merely puzzling, it seems patently irrational. For since this increase is entirely predictable, surely you could be made to lose money in a sequence of bets. At 6 PM you will be willing to accept a bet on heads at even odds, and at 11:59 PM you will, almost certainly, be willing to accept a bet on tails at worse than even odds. And that adds up to a sure loss. And surely that means you are irrational.

Now, one might think that this last argument shows that your degree of belief in tails in fact should not go up prior to midnight. One might indeed claim that since your degree of belief in heads should remain ½ until midnight, you should adjust your idea of what time it is when you see that the light is still on, rather than adjust your degree of belief in tails as time passes. But of course, this suggestion is impossible to carry out. Armed with an imperfect internal clock, you simply cannot make sure that your degree of belief in heads stays ½ until midnight, while allowing it to go down after midnight. So how should they develop?

Let us start with a much simpler case. Let us suppose that there is no coin toss and no light switching (and that you know this). You go into your cell at 6 PM. As time goes by there will be some development of your degrees of belief as to what time it is. Let us suppose that your degrees of belief in possible times develop as pictured in the top half of Figure 1.

Next, let us ask how your degrees of belief should develop were you to know with certainty that the guard will switch the light off at midnight, 12 PM. It should be clear then that at 11:59 PM your degree of belief distribution should be entirely confined to the left of midnight, as depicted in the bottom half of Figure 1. For at 11:59 PM the light will still be on, so that you know that it must be before 12 PM. But other than that it should be entirely confined to the left of 12 PM, it is not immediately clear exactly what your degree of belief distribution should be at 11:59 PM. It is not even obvious that there should be a unique answer to this question. A very simple consideration, however, leads to a unique answer.

Suppose that, even though the guard is going to switch the light off at 12 PM, you were not told that the guard is going to switch the light off at 12 PM. Then the development of your degree of belief would be as pictured in the top half of Figure 1. Next, suppose that

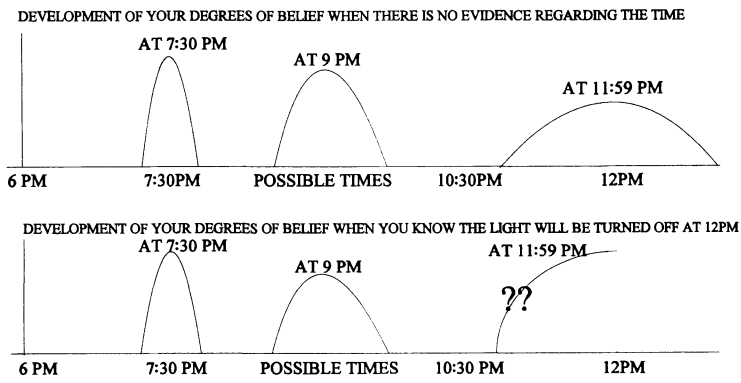

Figure 1

at 11:59 PM you are told that the guard will switch the light off at 12 PM, but you are not told that it is now 11:59 PM. Obviously, since the light is still on you can infer that it is prior to 12 PM. Surely you should update your degrees of belief by conditionalization: you should erase that part of your degree of belief distribution that is to the right of 12 PM, and renormalize the remaining part (increase the remaining part proportionally). Now it is clear that this is also the degree of belief distribution that you should have arrived at had you known all along that the guard would turn the light off at 12 PM. For either way you have accumulated exactly the same relevant information and experience by 11:59 PM. This uniquely determines how your degree of belief distribution should develop when you know all along that the guard will turn the light off at 12 PM. At any time this (constrained) distribution should be the distribution that you arrive at by conditionalizing the distribution that you have if you have no evidence regarding the time, on the fact that it is now before 12 PM. One can picture this development in the following way. One takes the development of the top part of Figure 1. As this distribution starts to pass through the 12 PM boundary, the part that passes through this boundary gets erased, and, in a continuous manner, it gets proportionally added to the part that is to the left of the 12 PM boundary.

Now we are ready to solve the original puzzle. Your degrees of belief in that case can be pictured as being distributed over possible times in two possible worlds: see Figure 2. The development is now such that when the bottom part of the degree of belief distribution hits midnight, it gets snuffed out to the right of midnight, and the rest of the degree of belief distribution is continuously renormalized, that is, the top part of the degree of belief distribution and the

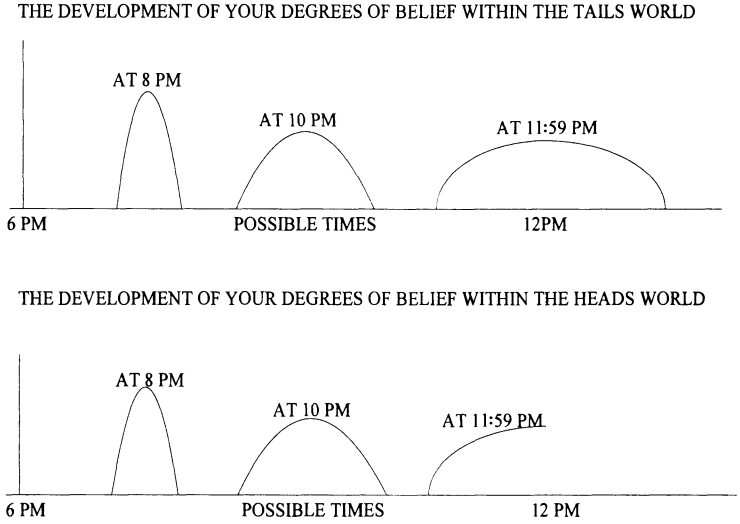

Figure 2

remaining bottom part are continuously proportionally increased as time passes. Note that Figure 2 is essentially different from Figure 1. In Figure 2, the top distribution starts to increase its absolute size once the leading edge of the bottom distribution hits midnight. This does not happen in Figure 1, since there the degree of belief distributions each were total degree of belief distributions in separate scenarios. Also, in Figure 2, the bottom distribution starts to increase in size once its leading edge hits midnight, but it only increases half as much as it does in Figure 1, since half of the "gains" is being diverted to the top degree of belief distribution.

Thus, at the very least, until it actually is midnight, the top and the bottom degrees of belief distribution will always be identical to each other, in terms of shape and size, to the left of midnight. Prior to midnight, your degrees of belief will be such that conditional upon it being prior to midnight, it is equally likely to be heads as tails. Your unconditional degree of belief in tails, however, will increase monotonically as you approach midnight.

After midnight there are two possible ways in which your degree of belief distribution can develop. If the light is switched off, your degree of belief distribution collapses completely onto midnight and onto the heads world. If in fact it is not switched off, your degree of belief distribution continues to move to the right in both worlds, and it continues to be snuffed out in the heads world to the right of

midnight, and the remaining degrees of belief keep being proportionally increased.[1]

Now I can answer the questions that I started with. It is true, as I surmised, that your degree of belief in tails will have increased by 11:59 PM. You will take your internal sense of the passing of time, and combine it with the fact that the light is still on, and you will take this as providing some evidence that the outcome is tails. It is also true, as I surmised, that the light still being on will be taken by you as providing some evidence that it is not yet midnight. For at 11:59 PM your degree of belief distribution over possible times (averaged over the heads and tails worlds) will be further to the left than it would have been had you believed that the light would stay on no matter what. More generally, we have found a unique solution to the puzzle of how a rational person's sense of time must interact with evidence, given how that person's sense of time works in the absence of evidence.

Rather surprisingly, this interaction can be such, as it is in my example, that you know in advance that at some specified later time you will, almost certainly, have increased your degree of belief in tails, and that you could not possibly have decreased your degree of belief in tails.[2] It is also interesting to note that nothing essential changes in this example if one assumes that the coin toss will take place exactly at midnight. Thus it can be the case that one knows in advance that one will increase one's degrees of belief that a coin toss, *which is yet*

[1] Thus, for instance, if the light is not switched off, there must be a moment (which could be before or after midnight) such that you have an equal degree of belief in each of the three possibilities: heads and it is before midnight, tails and it is before midnight, tails and it is after midnight.

[2] One might wonder why I inserted the phrase 'almost certainly' in this sentence. The reason for this is that there is a subtlety as to whether you know at 6 PM that you will have an increased degree of belief in tails at 11:59 PM. There is an incoherence in assuming that at 6 PM you know with certainty what your degree of belief distribution over possible times will be at 11:59 PM. For if you knew that, you could simply wait until your degree of belief distribution was exactly like that. (You can presumably establish by introspection what your degree of belief distribution is.) And when you reach that distribution, you would know that it has to be 11:59 PM. So when that happens you should then collapse your degree of belief distribution completely on it being 11:59 PM. But this is incoherent. Thus, the fact that you do not have a perfect internal clock also implies that you cannot know in advance what your degree of belief distribution is going to look like after it has developed (guided only by your internal clock). Thus you cannot in advance be certain how your degree of belief distribution over possible times will develop. Nonetheless, you can be certain at 6 PM that your degree of belief in tails will not decrease prior to midnight, and that it is extremely likely to have increased by 11:59 PM. At 6 PM your expectation for your degree of belief in tails at 11:59 PM will be substantially greater than ½.

to occur, will land tails. Of course, at the time that one has this increased degree of belief one does not know that this coin toss is yet to occur. Nonetheless, such predictable increases in degrees of belief seem very strange.

III. JOHN COLLINS'S PRISONER

John Collins has come up with the following variation of the case of the prisoner that was described in the previous section. In Collins's variation, the prisoner has two clocks in his cell, both of which run perfectly accurately. However, clock A initially reads 6 PM, clock B initially reads 7 PM. The prisoner knows that one of the clocks is set accurately, the other one is one hour off. The prisoner has no idea which one is set accurately; indeed, he initially has degree of belief $\frac{1}{2}$ that A is set accurately, and degree of belief $\frac{1}{2}$ that B is set accurately. As in the original case, if the coin lands heads the light in his cell will be turned off at midnight, and it will stay on if it lands tails. So initially the prisoner has degree of belief $\frac{1}{4}$ in each of the following four possible worlds:

W_1: heads and clock A is correct
W_2: heads and clock B is correct
W_3: tails and clock A is correct
W_4: tails and clock B is correct.

When in fact it is 11:30 PM the light, for sure, will still be on. What will the prisoner's degrees of belief then be? Well, if the actual world is W_1, then, when it actually is 11:30 PM clock A will read 11:30 PM and clock B will read 12:30 AM. In that case, since the prisoner sees that the light is still on, he will know that it cannot be that the coin landed heads and clock B is correct. That is to say, his degree of belief in W_2 will be 0, and his degrees of belief in the three remaining options will be $\frac{1}{3}$ each. Similarly if the actual world is W_3 then at 11:30 PM the prisoner will have degree of belief 0 in W_2 and degree of belief $\frac{1}{3}$ in each of the remaining options. On the other hand, if the actual world is W_2 or W_4, then when it is actually 11:30 PM, the clock readings will be 10:30 PM and 11:30 PM, and the prisoner will still have the degrees of belief that he started with, namely $\frac{1}{4}$ in each of the four possibilities. The prisoner, moreover, knows all of this in advance.

This is rather bizarre, to say the least. For, in the first place, at 6 PM the prisoner knows that at 11:30 PM his degrees of belief in heads will be less or equal to what they now are, and cannot be greater. So his current expectation of what his degrees of belief in heads will be at 11:30 PM is less than his current degree of belief in heads. Second, there is a clear sense in which he does not trust his future degrees

of belief, even though he does not think that he is, or will be, irrational, and even though he can acquire new evidence (the light being on or off). Let D_t denote the prisoner's degrees of belief at time t. Then, for example, $D_{6:00}$(clock B is correct/ $D_{11:30}$ (clock B is correct) = $\frac{1}{3}$) = 0. For $D_{11:30}$ (clock B is correct) = $\frac{1}{3}$ only occurs in worlds W_1 and W_3, and in each of those worlds clock B is not correct, and the prisoner knows this. Thus his current degrees of belief conditional upon his future degrees of belief do not equal those future degrees of belief. So he systematically distrusts his future degrees of belief. Strange indeed.

IV. SLEEPING BEAUTY

Some researchers are going to put Sleeping Beauty to sleep on Sunday night. During the two days that her sleep will last the researchers will wake her up either once, on Monday morning, or twice, on Monday morning and Tuesday morning. They will toss a fair coin Sunday night in order to determine whether she will be woken up once or twice: if it lands heads she will be woken upon Monday only, if it lands tails she will be woken up on Monday and Tuesday. After each waking, she will be asked what her degree of belief is that the outcome of the coin toss is heads. After she has given her answer she will be given a drug that erases her memory of the waking up; indeed it resets her mental state to the state that it was on Sunday just before she was put to sleep. Then she is put to sleep again. The question now is: When she wakes up, what should her degree of belief be that the outcome was heads?

> *Answer 1*: Her degree of belief in heads should be $\frac{1}{2}$. It was a fair coin and she learned nothing relevant by waking up.
>
> *Answer 2*: Her degree of belief in heads should be $\frac{1}{3}$. If this experiment is repeated many times, approximately $\frac{1}{3}$ of the awakenings will be heads-awakenings—that is, awakenings that happen on trials in which the coin landed heads.

Adam Elga[3] has argued for the second answer. I agree with him, and I agree with his argument. But let me amplify this view by giving a different argument for the same conclusion. Suppose that Sleeping Beauty is a frequent and rational dreamer. Suppose in fact that every morning if Sleeping Beauty is not woken up at 9 AM, she dreams at 9 AM that she is woken up at 9 AM. Suppose that the dream and reality are indistinguishable in terms of her experience, except that

[3] "Self-Locating Belief and the Sleeping Beauty Problem," *Analysis*, LX (2000): 143–47.

if Sleeping Beauty pinches herself and she is dreaming, it does not hurt (and she does not wake up), while if she does this while she is awake it does hurt. And let us suppose that Sleeping Beauty always remembers to pinch herself a few minutes after she experiences waking up (whether for real, or in a dream.) What should her degrees of belief be when she experiences waking up? It seems obvious she should consider the four possibilities equally likely (the four possibilities being: Monday & Tails & Awake, Monday & Heads & Awake, Tuesday & Tails & Awake, Tuesday & Heads & Dreaming). If Sleeping Beauty then pinches herself and finds herself to be awake, she should conditionalize and then have degree of belief $\frac{1}{3}$ in each of the remaining three possibilities (Monday & Tails & Awake, Monday & Heads & Awake, Tuesday & Tails & Awake). Suppose now that at some point in her life Sleeping Beauty loses the habit of dreaming. She no longer needs to pinch herself; directly upon waking she knows that she is not asleep. It seems clear, however, that this lack of dreaming should make no difference as to her degrees of belief upon realizing that she is awake. The process now occurs immediately, without the need for a pinch, but the end result ought to be the same.

Here again, the crucial assumption is commutativity: if the relevant evidence and experience collected is the same, then the order of collection should not matter for the final degrees of belief.[4] But there is clearly something very puzzling about such foreseeable changes in degrees of belief.

V. DUPLICATION

Scenario 1: While you are at the beach, Vishnu tells you that, contrary to appearances, you have existed only for one month: Brahma created you one month ago, complete with all your memories, habits, bad back, and everything. What is more, says Vishnu, one month ago Brahma in fact created two human beings like you (you are one of them), in exactly the same environment, at two different ends of the universe: one on earth, one on twin earth. Unfortunately, Vishnu has a further surprise for you: one month ago Shiva tossed a coin. If it landed heads, Shiva will destroy the human being that is on twin earth one month from now. If it landed tails, Shiva will do nothing. Vishnu does not tell you whether you are to be destroyed, but recommends that if you want to know, you should go check your mail at

[4] Cian Dorr has independently arrived at the idea of using commutativity in order to argue for the degrees of belief that Elga advocates in the Sleeping Beauty case—see Dorr, "Sleeping Beauty: In Defense of Elga," *Analysis* (forthcoming).

home. If there is a letter from President Bush for you, then you will be destroyed. Before running home, what degree of belief should you have in the four possibilities (Earth & Heads, Earth & Tails, Twin Earth & Heads, Twin Earth & Tails)? It seems clear that you should have degree of belief ¼ in each, or at the very least, that it is not irrational to have degree of belief ¼ in each. You run home, and find no letter from Bush. What should your degrees of belief now be? Well, by conditionalization, they should now be ⅓ in each of the remaining possibilities (Earth & Tails, Twin Earth & Heads, Twin Earth & Tails). Consequently you should now have degree of belief ⅓ that the toss landed heads and ⅔ that it landed tails.

Scenario 2: same as scenario 1, except that Vishnu tells you that if the toss came heads, your identical twin was destroyed by Shiva one week ago. Since you were obviously not destroyed, you do not need to rush home to look for a letter from Bush. In essence, you have learned the same as you learned in the previous scenario when you found you had no letter from Bush, and hence you should now have degree of belief ⅓ that the toss landed heads.

Scenario 3: same as scenario 2, except that Vishnu tells you that rather than that two beings were created one month ago by Brahma, one of them already existed and had exactly the life you remember having had. This makes no relevant difference and you should now have degree of belief ⅓ that the coin landed heads.

Scenario 4: same as scenario 3, except that Vishnu tells you that if the coin landed heads one month ago, Shiva immediately prevented Brahma from creating the additional human being one month ago. The upshot is that only if the coin landed tails will Brahma have created the additional human being. Since the timing of the destruction/prevention makes no relevant difference, you should again have degree of belief ⅓ that the coin landed heads.

Scenario 5:[5] you are on earth, and you know it. Vishnu tells you that one month from now Brahma will toss a coin. If it lands tails, Brahma will create, at the other end of the universe, another human being identical to you, in the same state as you will then be, and in an identical environment as you will then be. What do you now think that your degrees of belief should be in one month's time? The answer is that they should be the same as they are in scenario 5, since in one month's time you will be in exactly the epistemic situation that is described in scenario 5. Of course, it is plausible to claim that your

[5] This scenario is similar to the "Dr. Evil scenario" in Elga, "Defeating Dr. Evil with Self-Locating Belief" (manuscript).

future self will actually be on earth, since it is only your future continuation on earth that can plausibly be called "your future self." That does not mean, however, that your future self can be sure that he is on earth. For your future self will know that he will have the same experiences and memories, whether or not he is on earth or on twin earth, and thus he will not know whether he can trust his memories. Thus you now have degree of belief ½ in heads, and yet you know that in one month's time, you will have degree of belief ⅓. This is bizarre, to say the least.

Yet again, the crucial assumption in this reasoning is commutativity: your final degrees of belief should not depend on the order in which you receive all the relevant experience and evidence. You should end up with the same degree of belief—namely, degree of belief ½ in heads, whether you all along knew you were on earth, or whether you only later found out that you were on earth. But that can only be so if you had degree of belief ⅓ in heads prior to discovering that you were on earth.

VI. DIAGNOSIS

van Fraassen's reflection principle[6] says that one should trust one's future degrees of belief in the sense that one's current degree of belief D_0 in any proposition X, given that one's future degree of belief D_t in X equals p, should be p: $D_0(X/D_t(X)=p)=p$. Given that one is sure that one will have precise degrees of belief at time t, the reflection principle entails that one's current degrees of belief equal the expectations of one's future degrees of belief: $D_0(X)=\Sigma pD_0(D_t(X)=p)$. The reflection principle is violated in each of the five puzzles that I have presented, for in each case there is a time at which one's expectation of one's future degree of belief in heads differs from one's current degree of belief in heads. This is presumably why we find these cases, prima facie, so worrying and strange.

The source of the problem, I claim, is that the degrees of belief of perfectly rational people, people who are not subject to memory loss or any other cognitive defect, can develop in ways that are as yet unrecognized, and indeed are not allowed according to standard Bayesian lore. Standard Bayesian lore has it that rational people satisfy the principle of conditionalization: rational people alter their degrees of belief only by strict conditionalization on the evidence that they

[6] See van Fraassen, "Belief and the Problem of Ulysses and the Sirens," *Philosophical Studies*, LXXVII (1995): 7–37.

acquire.[7] Strict conditionalization of one's degrees of belief upon proposition X can be pictured in the following manner. One's degrees of belief are a function on the set of possibilities that one entertains. Since this function satisfies the axioms of probability theory it is normalized: it integrates (over all possibilities) to one. Conditionalizing such a function on proposition X then amounts to the following: the function is set to zero over those possibilities that are inconsistent with X, while the remaining nonzero part of the function is boosted (by the same factor) everywhere so that it integrates to one once again. Thus, without being too rigorous about it, it is clear that conditionalization can only serve to "narrow down" one's degree of belief distribution (one really *learns* by conditionalization). In particular a degree of belief distribution that becomes more "spread out" as time passes cannot be developing by conditionalization, and a degree of belief distribution that exactly retains its shape, but is shifted as a whole over the space of possibilities, cannot be developing by conditionalization. Such spreading out and shifting, however, is exactly what occurs in the five puzzles that I presented.

The reasons for such spreading and shifting are very simple. First, let us consider shifting. Suppose that one knows exactly what the history of the world that one inhabits is like. And suppose that one is constantly looking at a clock one knows to be perfect. One's degrees of belief will then be entirely concentrated on one possible world, and at any given moment one's degrees of belief within that world will be entirely concentrated on one temporal location, namely, the one that corresponds to the clock reading that one is then seeing. And that of course means that the location where one's degree of belief distribution is concentrated is constantly moving. That is to say, one's degree of belief distribution is constantly shifting, and such a constant shifting is simply not a case of conditionalization. Self-locating beliefs will therefore generically develop in ways that violate conditionalization. Collins's prisoner case involves exactly such a shifting of one's self-locating degrees of belief. The only difference is that, in his case, one additionally has an initial uncertainty as to which clock is accurate, that is, one is initially uncertain whether one is in a world in which clock A is correct or one in which clock B is correct. It is somewhat surprising that this kind of violation of conditionalization

[7] Strict conditionalization: when one learns proposition X at t, one's new degree of belief D_t equals one's old degree of belief D_0 conditional upon X : $D_t(Y)=D_0(Y/X)$. One might also allow Jeffrey conditionalization. It matters not for our purposes.

can be parlayed into a violation of reflection. But Collins's prisoner case shows exactly how one can do this.

Next, let us consider spreading. The simplest case of spreading is the case of the traveler who takes the path by the Mountains to Shangri La. His degrees of belief become more spread out when he arrives in Shangri La: at that time he goes from degrees of belief one in heads and zero in tails, to degrees of belief ½ in heads and ½ in tails.[8] The reason why this happens is that there are two distinct possible experiential paths that end up in the same experiential state. That is to say, the traveler's experiences earlier on determine whether possibility *A* is the case (Path by the Mountain), or whether possibility *B* is the case (Path by the Ocean). But because of the memory replacement that occurs if possibility *B* is the case, those different experiential paths merge into the same experience, so that that experience is not sufficient to tell which path was taken. Our traveler therefore has an unfortunate loss of information, due to the loss of the discriminating power of his experience. What is somewhat surprising is that this loss of discriminating power is not due to any loss of memory or any cognitive defect on his part: it is due to the fact that something strange would have happened to him had he taken the other path! This loss of discriminatory power of experience, and consequent spreading out of degrees of belief here does not involve self-locating degrees of belief. Suppose, for example, that our traveler is the only person ever to travel along either path. Then our traveler initially is unsure whether he is in a world in which path *A* is never taken or whether he is in a world in which path *B* is never taken. He then becomes sure that he is in a world in which path *B* is never taken. Even later,

[8] van Fraassen, in conversation with me, has suggested that in such situations conditionalization indeed should be violated, but reflection should not. In particular, he suggested that the degrees of belief of the traveler should become *completely vague*, upon arrival in Shangri La. This does not strike me as plausible. Surely upon arrival in Shangri La our traveler is effectively in the same epistemic situation as someone who simply knows that a fair coin has been tossed. One can make this vivid by considering two travelers, *A* and *B*. Traveler *A* never looks out of the window of the car, and hence maintains degree of belief ½ in heads all the way. (The memory replacement device does not operate on travelers who never look out of the window.) Traveler *A*, even by van Fraassen's lights, upon arrival in Shangri La, should still have degree of belief ½ in heads. Traveler *B*, however, does look out of the window during the trip. Upon arrival, by van Fraassen's lights, *B*'s degrees of belief should become completely vague. But it seems odd to me that traveler *B* is epistemically penalized, that is, is forced to acquire completely vague degrees of belief, just because he looked out of the window during the trip, when it seems clear that he ends up in exactly the same epistemic position as his companion, who did not look out of the window.

upon arrival, he again becomes unsure as to which world he is in. None of this has anything to do with self-locating beliefs.[9]

The source of the Sleeping Beauty and Duplication problems is exactly the same. In the case of Sleeping Beauty, the possibility of memory erasure ensures that the self-locating degrees of belief of Sleeping Beauty, even on Monday when she has suffered no memory erasure, become spread out over two days. In the Duplication case, yet again, the possible duplication of experiences forces one to become uncertain as to where (or who) one is. The cause of the spreading of degrees of belief in both cases is "experience duplication," and has nothing to do with the self-locating nature of these beliefs.[10]

It is not very surprising that the spreading of degrees of belief can bring about a violation of reflection. For instance, in the non-self-locating case a predictable reduction from degree of belief one in some proposition X to anything less than one will immediately violate reflection: now you know it, now you do not. The argument is slightly less straightforward in the self-locating case. Consider, for example, a case in which one is on Earth and one knows that at midnight a duplicate of oneself will be created on Mars. One might claim that since one now is certain that one is on Earth, and at midnight one will be uncertain as to whether one is on Earth, thus one has a clear violation of reflection. This is too quick, however. To have a clear violation of reflection it has to be the very same "content of belief" such that one's current degree of belief differs from one's expectation of one's future degree of belief. Depending on what one takes to be the contents of belief when it concerns self-locating beliefs (propositions? maps from locations to propositions?...), one might argue that the

[9] It is obvious how to generalize this case to a case in which there are memory replacement devices at the end of both roads, where these memory replacement devices are indeterministic, that is, when it is the case that for each possible path there are certain objective chances for certain memories upon arrival in Shangri La. For, given such chances (and the principal principle), one can easily calculate the degrees of belief that one should have (in heads and tails) given the memory state that one ends up with. And, generically, one will still violate conditionalization and reflection.

[10] Some people will balk at some of the degrees of belief that I have argued for in this paper, in particular in the self-locating cases. For instance, some people will insist that tomorrow one should still be certain that one is on Earth, even when one now knows (for sure) that a perfect duplicate of oneself will be created on Mars at midnight tonight. I beg to differ. Even if in this case, and other cases, however, one disagrees with me as to which degrees of belief are rationally mandated, the main claim of this paper still stands. The main claim is that in such cases of possible experience duplication, it is at the very least *rationally permissible* that one's degrees of belief become more spread out as time progresses, and hence rational people can violate conditionalization and reflection.

contents of belief are not the same at the two different times, and hence there is no violation of reflection. The arguments of sections IV and V, however, show that one can in any case parlay such spreading of self-locating degrees of belief into violations of reflection concerning such ordinary beliefs as to whether a coin lands heads or tails. So reflection is suckered anyhow.

Finally, the original case of the prisoner involves both a spreading of degrees of belief and a shifting of degrees of belief. The shifting is due simply to the passage of time and the self-locating nature of the beliefs. The spreading is due to the fact that our prisoner does not have experiences that are discriminating enough to pick out a unique location in time.[11] The analysis of section II shows, yet again, that such a spreading and shifting of self-locating degrees of belief can be parlayed into a violation of reflection concerning such ordinary beliefs as to whether a coin lands heads or tails.

VII. CONCLUSIONS

The degrees of belief of rational people can undergo two as yet unrecognized types of development. Such degrees of belief can become more spread out due to the duplication of experiences, or more generally, due to the loss of discriminating power of experiences, and thereby violate conditionalization. In addition, self-locating degrees of belief will generically be shifted over the space of possible locations, due to the passage of time, and thereby violate conditionalization. Such violations of conditionalization can be parlayed into violations of reflection, and lead to a distrust of one's future degrees of belief. Strange, but not irrational.

FRANK ARNTZENIUS

Rutgers University

[11] One might model the prisoner here as having unique distinct experiences at each distinct, external clock, time, and as initially having precise degrees of belief over the possible ways in which those experiences could correlate to the actual, external clock, time. If one were to do so, then the prisoner would merely be initially uncertain as to which world he was in (where worlds are distinguished by how his experiences line up with the actual, external clock, time), but for each such possible world would be always certain as to where he was located in it. And, if one were to do so, then the original prisoner case would be essentially the same case as Collins's prisoner case: no uncertainty of location in any given world, merely an initial uncertainty as to which world one is in, and a subsequent shifting of the locally concentrated degrees of belief within each of the possible worlds. There is no need, however, to represent the original prisoner case that way. Indeed, it seems psychologically somewhat implausible to do so. More importantly, the arguments and conclusions here do not depend on how one models this case.