Rescuing Reflection

Ilho Park*†

In this article, I suggest an argument that seems to show a conflict between the reflection principle and conditionalization. In particular, I show that when the reflection principle is formulated in a standard way, the principle conflicts with Jeffrey conditionalization. And it is argued that the source of the conflict resides in an ambiguity of the standard formulation. Furthermore, I attempt to rescue the principle using Bayes factors. That is, I suggest a new formulation of the principle so as to avoid the conflict.

1. Introduction. Probabilists think that our beliefs come in degrees. And they have suggested some constraints on our rational degrees of belief (i.e., credences and their rational updates). For example, our credences at a given time are required to be probabilistically coherent in the sense that they satisfy the axioms of probability theory, that is, nonnegativity, normalization, and (countable) additivity. It is also widely accepted that our credences should be updated using conditionalization. Here are two versions of conditionalization, that is, strict conditionalization (SC) and Jeffrey conditionalization (JC):

> **SC**: When at time $t + 1$ an observation leads an agent to be certain of E and nothing else, for any X, $P_{t+1}(X) = P_t(X|E)$, when $0 < P_t(E) < 1$.

> **JC**: When at time $t + 1$ an observation leads an agent to change directly her credence in E from $P_t(E)$ to $P_{t+1}(E)$ (<1) and nothing else is changed directly, for any X, $P_{t+1}(X) = P_{t+1}(E)P_t(X|E) + P_{t+1}(\sim E)P_t(X|\sim E)$, when $0 < P_t(E) < 1$.

Received September 2011; revised March 2012.

*To contact the author, please write to: Seoul University, 1 Gwanak-ro, Gwanak-gu, Seoul 151-742, Korea; e-mail: ipark.phil@gmail.com.

†I would like to thank the anonymous referees for their invaluable suggestions. And I owe special debts to Inkyo Chung and Alan Hájek for their detailed and insightful comments.

Philosophy of Science, 79 (October 2012) pp. 473–489. 0031-8248/2012/7904-0009$10.00
Copyright 2012 by the Philosophy of Science Association. All rights reserved.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

Here P_t and P_{t+1} are the agent's credence functions at time t and at time $t + 1$, respectively.

It is noteworthy that probabilistically coherent credence functions are still probabilistically coherent after updating by conditionalization. In this sense, we may say that these two constraints do not conflict with each other, or that probabilistic coherence of credence functions is preserved through updating by conditionalization. This relationship between two constraints is desirable. And if there are other rationality constraints, then they should not conflict with the above two constraints. In particular, any other synchronic constraints like probabilistic coherence should be satisfied after updating by conditionalization if they are satisfied before the updating. That is, preservation is natural:

> **Preservation**: For any rationality constraint C, if C holds for a credence function P, then C should also hold for a credence function P' that is obtained from P by conditionalization.

When a constraint satisfies preservation, I say that the constraint is preserved under conditionalization. In what follows, I suggest an argument that seems to show that another well-known rationality constraint conflicts with conditionalization (i.e., the constraint is not preserved under conditionalization).

The constraint in question is the *reflection principle*, formulated and named by van Fraassen (1984, 244). This is a constraint on our higher-order credences (i.e., credences in credences). This principle may be formulated in various ways. Here is a standard formulation that is suggested by Maher (1993, 105):

> **RP**: For any X and any real number $a > 0$, $P_t(X|R_{t+a}) = P_{t+a}(X)$, when $P_t(R_{t+a}) > 0$.

Here R_{t+a} refers to the proposition that the credence function is P_{t+a} at time $t+a$. Van Fraassen (1984, 244) originally formulates the principle in a slightly different way. Let P_{t+a} assign x to X at time $t+a$. Here is van Fraassen's formulation:

> **RP′**: For any X, real numbers x and a (>0) such that x is assigned to the credence in X at time $t + a$, $P_t(X|P_{t+a}(X) = x) = x$ when $P_t(P_{t+a}(X) = x) > 0$.

RP is a formulation that simplifies RP′ without the variable x and the phrase "x is assigned to the credence in X at time $t+a$." The following discussion does not depend on whether we adopt RP or RP′ as a formulation of the principle. For simplicity, I adopt RP rather than RP′ in what follows.

Informally, RP explicates a rational relationship between an agent's current and future credence functions. Why is it rational to obey RP? An intu-

https://doi.org/10.1086/668001 Published online by Cambridge University Press

itive reason is that our future selves are epistemically superior to our current selves, and it is rational to follow opinions of agents who are superior to us. Then, does RP not conflict with conditionalization? To put it another way, is RP preserved under conditionalization? More precisely, do agents who satisfied RP before updating by conditionalization still satisfy RP after the updating? In section 2, I suggest an argument that appears to show that RP conflicts with conditionalization—in particular JC. But I point out in section 3 that the conflict is due to an ambiguity of the standard formulation of the principle (i.e., RP). In section 4, furthermore, I attempt to reformulate the reflection principle so as to circumvent the prima facie conflict.

Before presenting my discussion, let me explain some expressions. If and only if an agent whose credence function is P_t satisfies RP, I say that P_t satisfies RP. That is, I use interchangeably 'an agent' and 'her credence function' in this article. And if and only if it holds that for any X, $P_t(X|R_{t+a}) = P_{t+a}(X)$ when $P_t(R_{t+a}) > 0$, I say that P_t satisfies RP concerning P_{t+a}.

2. A Conflict between Reflection and Conditionalization. I examine in this section whether RP is preserved under conditionalization. Consider figure 1, where ϵ stands for a course of experience, P is an agent's credence function at a given time, and Q is her credence function that is updated from P after the agent undergoes ϵ. It is well accepted by Bayesians that when a course of experience leads an agent to change her credence function, the new credence function is determined by her old credence function and the experience itself (see, e.g., Field 1978). Figure 1 is intended to describe this explicitly. That is, the figure represents that Q is determined by the old credence function P and the experience ϵ.

Now let's consider figure 2, where P_0 is an agent's credence function at time t, and P_1 and P_1' are credence functions that the agent might have at time $t + 1$. Similarly, P_2 and P_2' are credence functions that she might have at time $t + 2$. Meanwhile, ϵ_1 and ϵ_1' stand for possible courses of experience that the agent might undergo during $[t, t + 1]$, and ϵ_2 stands for a possible course of experience during $[t + 1, t + 2]$. Similar to figure 1, figure 2 describes what experience and what old credence function determine what new credence function. For example, P_1 is determined by P_0 and ϵ_1, and P_2 is determined by P_1 and ϵ_2. Since P_1 is determined by P_0 and ϵ_1, it can be said that P_2 is determined by P_0, ϵ_1, and ϵ_2.

Our purpose is to examine whether RP is preserved under conditionalization. Suppose that P_0 satisfies RP. Then, P_0 satisfies RP concerning P_2 as well as P_2'. And let's consider a case in which the agent undergoes ϵ_1 during $[t, t + 1]$, and so P_1 is obtained from P_0 by conditionalization. For our purpose, then, what should we do? I should emphasize here that in this case, P_1 does not have to satisfy RP concerning all future credence functions of P_0. In particular, P_1 does not have to satisfy RP concerning P_2'. This is be-

https://doi.org/10.1086/668001 Published online by Cambridge University Press

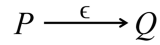


Figure 1.

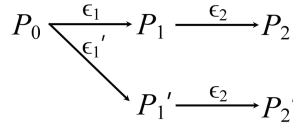


Figure 2.

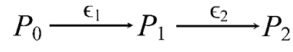


Figure 3.

cause P_2' is not a future credence function of P_1. If one credence function is a future credence function of another function, the former should be affected by all courses of experience that affect the latter. Note that P_2' is not affected by a course of experience that affects P_1—in particular, P_1 is affected by ϵ_1 but P_2' is not. Thus, P_2' is not a future credence function of P_1, and so we do not have to examine whether P_1 satisfies RP concerning P_2'. In this case, rather, all we should do is to examine whether P_1 satisfies RP concerning P_2.

With the help of this consideration, we can say that in order to examine whether RP is preserved under conditionalization, it is sufficient to consider the case represented by figure 3, where P_0 is an agent's credence function at a given time t, P_1 is updated from P_0 after the agent undergoes ϵ_1 at time $t + 1$, and P_2 is obtained from P_1 after the agent undergoes ϵ_2 at time $t + 2$. Then, all we have to do is to answer the following question: Does P_1 satisfy RP concerning P_2 if P_0 satisfies RP concerning P_2 and P_1 is obtained from P_0 by conditionalization? In response to this question, if an affirmative answer can be given, then it should be concluded that RP is preserved under conditionalization. Otherwise, we should say that RP is not preserved. Before seeking the answer, it should be emphasized that in this article, I rule out cases in which the future credence functions are affected by cognitive malfunction, memory loss, and so forth. Thus, it is assumed that an agent's future credence functions have all evidence that her current function has. In particular, it is assumed that P_2 has all evidence that P_1 has.

Now, it is time to examine whether RP is preserved under conditionalization. Let's begin with a simple case. Suppose that ϵ_1 leads the agent to be certain of E_1 at time $t + 1$. Thus, $P_1(E_1) = 1$. Here we should note that

https://doi.org/10.1086/668001 Published online by Cambridge University Press

$$P_2(E_1) = 1. \tag{1}$$

This is because P_2 has all evidence that P_1 has. At time $t + 2$, thus, the agent has the evidence that the agent learns from ϵ_1 at time $t + 1$, and so equation (1) holds. Now, suppose that the agent updates her credences using conditionalization. In this case, the agent updates her credence using SC. Then, we have that[1]

$$\text{For any } X,\ P_1(X) = P_0(X|E_1). \tag{2}$$

Moreover, let's suppose that P_0 satisfies RP concerning P_2. Thus, we have that[2]

$$\text{For any } X \text{ and } Y,\ P_0(X|R_2\&Y) = P_2(X|Y). \tag{3}$$

Here, R_2 refers to the proposition that the credence function is P_2 at time $t + 2$. Then, it follows from equations (1), (2), and (3) that[3]

$$\text{For any } X,\ P_1(X|R_2) = P_2(X). \tag{4}$$

Thus, we can say that if P_0 satisfies RP concerning P_2 and P_1 is obtained from P_0 by SC on E_1, then P_1 also satisfies RP concerning P_2. Thus, we can conclude that RP is preserved under SC. This is a desirable result.

However, we should note that we have considered just a simple case in which experience leads the agent to be certain of a proposition. Then, what if experience leads the agent to change her credence in some propositions to a value that is less than 1? In this case, Bayesians advise us to update our credences using another version of conditionalization (i.e., JC). Can we show that RP is also preserved under JC? Unfortunately, it seems not.

Suppose that ϵ_1 leads the agent to change her credence in E_1 from $P_0(E_1)$ to $P_1(E_1)$ (<1) at time $t + 1$ and that $P_2(E_1)$ is in (0, 1). Then, we have that

$$P_0(E_1) \neq P_1(E_1). \tag{5}$$

Suppose also that the agent is a JC updater, that is, she updates her credence

1. Here, it is assumed that $P_0(E_1) > 0$, and so $P_0(X|E_1)$ is well defined.

2. Here, it is assumed that $P_0(R_2\&Y) > 0$, and so $P_0(X|R_2\&Y)$ is well defined. Now, suppose that P_0 satisfies RP concerning P_2. Then, we have that (*a*) for any X, $P_0(X|R_2) = P_2(X)$ when $P_0(R_2) > 0$. And the probability calculus says that for any X and Y, $P_0(X|R_2\&Y) = P_0(X\&Y|R_2)/P_0(Y|R_2)$ when $P_0(R_2\&Y) > 0$. From *a*, thus, it follows that for any X and Y, $P_0(X|R_2\&Y) = P_2(X\&Y)/P_2(Y) = P_2(X|Y)$ when $P_0(R_2\&Y) > 0$. Note that it is assumed that $P_0(R_2\&Y) > 0$. Thus, we have eq. (3).

3. From eq. (2), it follows that (*a*) for any X, $P_1(X|R_2) = P_0(X|R_2\&E_1)$. By eq. (3) and *a*, it is implied that (*b*) for any X, $P_1(X|R_2) = P_0(X|R_2\&E_1) = P_2(X|E_1)$. And eq. (1) says that $P_2(E_1) = 1$, and so $P_2(X|E_1) = P_2(X)$. Thus, it follows from *b* that for any X, $P_1(X|R_2) = P_2(X)$. Now, we have eq. (4).

https://doi.org/10.1086/668001 Published online by Cambridge University Press

function using JC. In particular, let's assume that P_1 is obtained from P_0 by JC on $\{E_1, \sim E_1\}$. Then, it follows that[4]

$$\begin{aligned} &\text{For any } X,\ P_0(X|E_1) = P_1(X|E_1), \text{ and} \\ &P_0(X|\sim E_1) = P_1(X|\sim E_1) \text{ when } 0 < P_0(E_1) < 1. \end{aligned} \tag{6}$$

Now, suppose that P_0 satisfies RP concerning P_2. Then, it holds that

$$\text{For any } X,\ P_0(X|R_2) = P_2(X) \text{ when } P_0(R_2) > 0. \tag{7}$$

Then, can it follow from equations (5), (6), and (7) that P_1 also satisfies RP concerning P_2? No. Suppose, for reductio, that P_1 also satisfies RP concerning P_2. Then, it holds that

$$\text{For any } X,\ P_1(X|R_2) = P_2(X) \text{ when } P_1(R_2) > 0. \tag{8}$$

Now we have the following three equations:[5]

$$\begin{aligned} P_0(R_2|E_1) &= P_1(R_2|E_1); \\ P_0(R_2|\sim E_1) &= P_1(R_2|\sim E_1); \\ P_0(E_1|R_2) &= P_1(E_1|R_2). \end{aligned}$$

And these equations imply that $P_0(E_1) = P_1(E_1)$, which contradicts equation (5) (see app. A). Thus, we should conclude that equations (5)–(8) jointly imply a contradiction, and so RP is not preserved under JC. In other words, even if an agent's credence function satisfies RP before updating using JC, she cannot satisfy RP after the updating. This is an undesirable result. If RP is a plausible rationality constraint, then it is natural that RP should satisfy preservation. But the above result says that RP does not. In particular, RP is not preserved under one version of conditionalization (i.e., JC).[6] How can we respond to this problem? In the next section, I try to answer this question.

4. It is well known that JC is equivalent with *rigidity*: when at time $t + 1$, an observation leads an agent to change directly her credence in E from $P_t(E)$ to $P_{t+1}(E)$ (<1) and nothing else is changed directly, for any X, $P_t(X|E) = P_{t+1}(X|E)$ and $P_t(X|\sim E) = P_{t+1}(X|\sim E)$, when $0 < P_t(E) < 1$.

5. Here, it is assumed that $0 < P_0(E_1) < 1$, $P_0(R_2) > 0$, and $P_1(R_2) > 0$. Thus, eq. (6) implies that $P_0(R_2|E_i) = P_1(R_2|E_1)$ and $P_0(R_2|\sim E_1) = P_i(R_2|\sim E_1)$. And it follows from eqq. (7) and (8) that $P_0(E_1|R_2) = P_1(E_1|R_2)$.

6. Some readers may be curious as to why RP is preserved under SC but not JC. They should pay attention to the fact that when experience leads an agent to be certain of a proposition at a given time, the agent's future credence functions assign 1 to the proposition. So, eq. (1) holds. However, when experience leads an agent to change her credence in a proposition to a value less than 1, the agent's future credence functions do not have to assign a particular value to the proposition. Thus, eq. (1) does not have to hold in such cases. And this difference leads to the result that RP is preserved under SC but not JC.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

3. Diagnosis: An Ambiguity of RP. In this section, I explain first why RP cannot be satisfied when a credence function is updated by JC on a partition. With the help of this explanation, I argue that RP is not a plausible formulation of the reflection principle.

Let's revisit figure 3, first. Suppose that P_0 satisfies RP concerning P_2. Then, we have that $P_0(X|R_2) = P_2(X)$, for any X. As was mentioned earlier, P_2 is determined by P_0, ϵ_1, and ϵ_2. So, we can say that the credence function $P_0(\cdot|R_2)$ is also determined by P_0, ϵ_1, and ϵ_2. Now, let's consider a case in which the agent actually undergoes ϵ_1 during $[t,\ t+1]$. In this case, $P_1(\cdot|R_2)$ is obtained from $P_0(\cdot|R_2)$ by JC. We can say, then, that $P_1(\cdot|R_2)$ is determined by the old credence function $P_0(\cdot|R_2)$ and the experience ϵ_1. In other words, $P_0(\cdot|R_2)$ and ϵ_1 jointly affect $P_1(\cdot|R_2)$ through the updating. It should be noted again that $P_0(\cdot|R_2)$ is affected by P_0, ϵ_1, and ϵ_2. Thus, the impact of ϵ_1 has already been reflected in $P_0(\cdot|R_2)$ before the updating. Nevertheless, if the experience ϵ_1 affects $P_0(\cdot|R_2)$ once again through the updating, then it should be said that the updated credence function, that is, $P_1(\cdot|R_2)$, is overdetermined by ϵ_1. However, P_2 is determined by P_0, ϵ_1, and ϵ_2. This credence function P_2 is not overdetermined by ϵ_1. That is, P_2 is not twice affected by ϵ_1. For this reason, $P_1(\cdot|R_2)$ that is obtained from $P_0(\cdot|R_2)$ by JC and so is overdetermined by ϵ_1 cannot be equal to P_2 that is not overdetermined by ϵ_1. Despite this fact, however, RP still requires that $P_1(X|R_2) = P_2(X)$, for any X. Hence, we cannot help concluding that RP conflicts with JC.

Then, should we abandon at least one of the reflection principle and JC? I think we do not have to. In particular, we do not have to give up the reflection principle. I think that what should be abandoned is not the reflection principle itself but the standard formulation of this principle (i.e., RP).[7] In what follows, I explain why RP is a misleading formulation.

Consider the proposition that the credence function is changed from P_t to P_{t+a}. (Here it is assumed that $a > 0$.) Let's abbreviate this proposition by using C_t^{t+a}. Then, what should be the value of $P_t(X|C_t^{t+a})$? RP says that for any X, $P_t(X|R_{t+a})$ should be equal to $P_{t+a}(X)$. Here, $P_t(X|R_{t+a})$ is the agent's conditional credence in X, given that her credence function is P_{t+a} at time $t + a$. Here, P_{t+a} is her future credence function that is updated from P_t. Thus, we can say that the information obtained from R_{t+a} in $P_t(X|R_{t+a})$ is as follows: (i) P_{t+a} is obtained from P_t; (ii) there is a course of experience that leads to changing the credence function from P_t to P_{t+a}.[8] Consider C_t^{t+a} in

7. As explained in n. 9, I show that van Fraassen's original formulation, i.e., RP′, has the same fault as RP.

8. In fact, R_{t+a} itself says nothing about from what credence function P_{t+a} is updated. However, R_{t+a} in the context of $P_t(X|R_{t+a})$ informs us that P_{t+a} is updated from P_t.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

$P_t(X|C_t^{t+a})$. It is intuitive, I think, that the information obtained from C_t^{t+a} is the same as the information from R_{t+a} in $P_t(X|R_{t+a})$. That is, C_t^{t+a} informs us of i and ii. Thus, it is natural that the value of $P_t(X|C_t^{t+a})$ is equal to the value of $P_t(X|R_{t+a})$. Hence, when we say that the reflection principle should be satisfied at time t, it is required that

$$\text{For any proposition } X \text{ and any real number } a > 0,$$
$$P_t(X|C_t^{t+a}) = P_{t+a}(X), \text{ when } P_t(C_t^{t+a}) > 0. \tag{9}$$

Now, let's revisit figure 3. Let C_0^2 be the proposition that the agent's credence function is changed from P_0 to P_2. Similarly, let C_1^2 be the proposition that the agent's credence function is changed from P_1 to P_2. As in the previous discussion, let's assume that the agent satisfies the reflection principle at time t. In particular, suppose that $P_0(X|C_0^2) = P_2(X)$, for any X. Suppose also that the agent undergoes ϵ_1 during $[t, t+1]$, and so P_0 is updated to P_1 by JC. Then, $P_0(X|C_0^2)$ is changed to $P_1(X|C_0^2)$. Recall that the argument in the previous section has shown that even if $P_0(X|R_2)$ is equal to $P_2(X)$, $P_1(X|R_2)$ obtained from $P_0(X|R_2)$ by JC cannot be equal to $P_2(X)$. Redescribing this result using 'C_0^2', it can be said that even if $P_0(X|C_0^2)$ is equal to $P_2(X)$, $P_1(X|C_0^2)$ obtained from $P_0(X|C_0^2)$ by JC cannot be equal to $P_2(X)$. Does it follow from this result that the reflection principle conflicts with JC? No. Here we should pay attention to what the principle requires in this case. According to equation (9), when we say that the reflection principle is satisfied at time $t+1$ what is required is not that

$$\text{For any } X,\ P_1(X|C_0^2) = P_2(X) \text{ when } P_1(C_0^2) > 0 \tag{10}$$

but that

$$\text{For any } X,\ P_1(X|C_1^2) = P_2(X) \text{ when } P_1(C_1^2) > 0. \tag{11}$$

For this reason, the result that $P_1(X|C_0^2)$ is not equal to $P_2(X)$ after the updating does not threaten the reflection principle and JC. In order to show that the reflection principle is not preserved under JC, we should demonstrate that equation (11), not (10), does not hold after JC updating. However, the argument in the previous section shows just that equation (10) does not hold after the updating.

To sum up, the source of the problem resides in the fact that the standard formulation of the reflection principle (i.e., RP) is ambiguous. In particular, when RP requires that $P_1(X|R_2) = P_2(X)$ for any X after the updating by JC, it is ambiguous whether the requirement means equation (10) or (11).[9] Now

9. Although we adopt van Fraassen's original formulation of the reflection principle, i.e., RP′, a similar ambiguity occurs. Suppose that an agent's future credence function P_2 as-

https://doi.org/10.1086/668001 Published online by Cambridge University Press

we can conclude that the prima facie conflict that is suggested in the previous section does not impugn the reflection principle and JC. The problem is resolved.

There is a limit to the above discussion, however. I have argued why the conflict between RP and JC cannot threaten the reflection principle itself and JC. But I have not shown positively that agents who satisfied the reflection principle before updating by JC also satisfy the principle after the updating. It is time to show this.

4. A New Formulation of the Reflection Principle. I have shown in section 3 that the conflict between RP and JC is due to an ambiguity of RP. In order to rescue the reflection principle from the conflict, thus, we should first reformulate the reflection principle so as to avoid this ambiguity. In this section, I suggest a new formulation of the principle and show that when the principle is formulated using the new one, the conflict in question disappears.

My suggestion is to reformulate the principle using C_t^{t+a}, rather than R_{t+a}. Suppose that P_t satisfies the reflection principle and that P_{t+1} is obtained from P_t by JC. Suppose also that the principle is preserved under JC. According to equation (9), then, the principle requires that

$$\text{For any } X \text{ and any real number } a > 1,$$
$$P_{t+1}(X|C_{t+1}^{t+a}) = P_{t+a}(X), \text{ when } P_{t+1}(C_{t+1}^{t+a}) > 0.$$

I should emphasize here that $P_{t+1}(\cdot|C_{t+1}^{t+a})$ is updated from $P_t(\cdot|C_{t+1}^{t+a})$—not $P_t(\cdot|C_t^{t+a})$—by JC. In other words, $P_{t+1}(\cdot|C_{t+1}^{t+a})$ is determined by the experience during $[t, t+1]$ and $P_t(\cdot|C_{t+1}^{t+a})$—not $P_t(\cdot|C_t^{t+a})$. Thus, we need to identify $P_t(\cdot|C_{t+1}^{t+a})$ in order to examine whether the reflection principle is preserved under JC.

Let's consider a conditional credence $P_t(X|C_{t+1}^{t+a})$. This credence is a conditional credence at t in X, given that the credence function is changed from P_{t+1} to P_{t+a}. The change from P_{t+1} to P_{t+a} is due to a course of experience during $[t+1, t+a]$. So, the condition could be naturally thought as saying that the agent undergoes the course of experience during $[t+1, t+a]$. But the reflection principle that hitherto has been discussed in this article says nothing about what the value of $P_t(X|C_{t+1}^{t+a})$ should be. As was shown, equation (9) just requires that $P_t(X|C_t^{t+a})$ should be equal to $P_{t+a}(X)$. More generally, equation (9) is silent on the value of $P_t(X|C_{t+b}^{t+a})$ when $0 < b < a$. For our pur-

signs c to X. And suppose that at time $t+1$, the agent updates her credence function from P_0 to P_1. Suppose even that P_0 and P_1 assign a and b to X, respectively. Then, RP′ requires that $P_1(X|P_2(X) = c)$ updated from $P_0(X|P_2(X) = c)$ should be equal to c. But, the expression '$P_1(X|P_2(X) = c)$' ambiguously stands for the following two conditional credences: $P_1(X|$the credence in X is changed from a to c at time $t+2)$; $P_1(X|$the credence in X is changed from b to c at time $t+2)$.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

pose, thus, the reflection principle should be formulated so that the value of $P_t(X|C_{t+b}^{t+a})$ can be determined.

As was mentioned, the reflection principle requires that $P_t(\cdot\,|C_t^{t+a})$ should be equal to P_{t+a}. Here, P_{t+a} is determined by P_t and a course of experience during $[t, t + a]$ that leads the credence function to be changed from P_t to P_{t+a}. Then, what function should $P_t(\cdot\,|C_{t+b}^{t+a})$ be equal to? It is natural that the credence function in question should be one that is determined by P_t and the course of experience during $[t + b, t + a]$ that leads the credence function to be changed from P_{t+b} to P_{t+a}. So, in order to determine the value of $P_t(X|C_{t+b}^{t+a})$, we need to find a way to combine the two factors—P_t and a course of experience during $[t + b, t + a]$—into a credence function. For this purpose, we should also be able to represent the course of experience during $[t + b, t + a]$ using a parameter.

Fortunately, there is a parameter that many Bayesians think plays such a role: the Bayes factors (see, e.g., Field 1978; Jeffrey 2002; Wagner 2002, 2003). Consider an outcome space $\Omega = \{\omega_i\}$ whose members are mutually exclusive and exhaustive. Suppose that there is an agent whose credence functions are measured on this space. Suppose also that her credence function is changed from P to Q. Then, the Bayes factor of ω_i against ω_1 is defined as follows:

$$\text{For any } \omega_i \in \Omega,\ \beta_P^Q(\omega_i, \omega_1) = \frac{Q(\omega_i)/P(\omega_i)}{Q(\omega_1)/P(\omega_1)}.$$

Here, ω_1 is an arbitrary anchored proposition that is a member of Ω. If we think that the Bayes factors represent the impact of experience that leads the agent's credence function to be changed from P to Q, then we can determine the value of $P_t(X|C_{t+b}^{t+a})$ using the Bayes factors in question and P_t. How can we do it?

This situation is very similar to the collaborative belief updating. Jeffrey (2002, 59–61) suggests a way to update our credences using another agent's observation. Suppose that an agent S_1 wants to update her credence function using her colleague S_2's observation. Suppose also that at time $t + 1$, the colleague S_2 observes something and changes her credence function from P_2 to Q_2. Then, how should S_1 use S_2's observation? Jeffrey says that S_1 cannot regard S_2's new credence function as her own new credence function since S_2's new credences are affected by S_2's old credences as well as the observation in question. Instead, he advises that S_1 should regard S_2's Bayes factors as her own Bayes factors.

Suppose that the credences of S_1 and of S_2 are measured on an outcome space $\Omega = \{\omega_i\}$. Let X be a disjunction of members of Ω. Suppose also that S_1's credence function is changed from P_1 to Q_1. Then, JC can be reformulated as follows:

https://doi.org/10.1086/668001 Published online by Cambridge University Press

$$\text{For any } X,\ Q_1(X) = \frac{\sum_i \beta_i^1 P_1(X \& \omega_i)}{\sum_i \beta_i^1 P_1(\omega_i)}.$$

Here β_i^1 is S_1's Bayes factor of ω_i against ω_1, that is, $\beta_{P_1}^{Q_1}(\omega_i,\ \omega_1)$. For the above collaborative belief updating, Jeffrey suggests replacing S_1's Bayes factor in the formula with S_2's Bayes factor $\beta_{P_2}^{Q_2}(\omega_i,\ \omega_1)$ that is obtained by her observation. Through this collaborative belief updating, thus, S_1's new credences are updated by the following rule:

$$\text{For any } X,\ Q_1(X) = \frac{\sum_i \beta_i^2 P_1(X \& \omega_i)}{\sum_i \beta_i^2 P_1(\omega_i)}. \tag{12}$$

Here, β_i^2 is S_2's Bayes factor of ω_i against ω_1, that is, $\beta_{P_2}^{Q_2}(\omega_i,\ \omega_1)$. Note that equation (12) describes a way to obtain the new credence function from the old credence function and the Bayes factors that represent experience itself. To put it another way, equation (12) suggests a way to combine the two factors—a credence function and a course of experience—into another credence function.

Recall that in order to determine the value of $P_t(X|C_{t+b}^{t+a})$, we have to seek a way to combine the two factors—P_t and a course of experience during $[t + b, t + a]$—into a credence function. Thus, a similar method to the collaborative belief updating can be applied. That is, we can determine the value of $P_t(X|C_{t+b}^{t+a})$ using P_t and the Bayes factors that are obtained from C_{t+b}^{t+a}, and so we are able to reformulate the reflection principle as we wish. Suppose that an agent's credence functions are measured on an outcome space $\Omega = \{\omega_i\}$. Here is my new version of the reflection principle:

NRP: For any proposition X and any real number $0 \le b < a$,

$$P_t(X|C_{t+b}^{t+a}) = \frac{\sum_i \beta_i P_t(X \& \omega_i)}{\sum_i \beta_i P_t(\omega_i)},\ \text{when } P_t(C_{t+b}^{t+a}) > 0.$$

Here, 'C_{t+b}^{t+a}' expresses that the credence function is changed from P_{t+b} to P_{t+a}. And β_i is the Bayes factor of ω_i against ω_1 that is obtained from C_{t+b}^{t+a}. That is,

$$\text{For any } \omega_i \in \Omega,\ \beta_i = \beta_{P_{t+b}}^{P_{t+a}}(\omega_i,\ \omega_1) = \frac{P_{t+a}(\omega_i)/P_{t+b}(\omega_i)}{P_{t+a}(\omega_1)/P_{t+b}(\omega_1)}.$$

https://doi.org/10.1086/668001 Published online by Cambridge University Press

Now, we should show that this NRP is preserved under conditionalization—in particular JC. Before presenting the proof, let's consider a relationship between RP and NRP.

What is the difference between RP and NRP? Suppose that P is your current credence function and you know this. Suppose also that you update your credences using conditionalization. And let's consider a case in which you get to know that Q is your future credence function. According to RP, then, your credence function should be changed to Q. Now, let's consider another case in which you get to know that you will undergo a course of experience that could change your credence function from P' to Q. Do you get to know that Q is your future credence function? It depends. If P' is the same as P, then you can get to know that Q is your future credence function. This is because you know that your current credence function is P and that the experience will lead your credence function to be changed from P to Q. In this case, RP requires that you should change your credence function to Q. However, if P' is not the same as P, then you cannot get to know that Q is your future credence function. All you get to know is just that you will undergo the experience in question—not that you will undergo the experience in question and so your credence function will be changed from P to Q. In this case, RP seems difficult to guide any change of your credence function. In other words, RP says nothing how the knowledge of your future experience should be reflected in your current credence function. It is natural, nevertheless, that the knowledge is informative and thus affects your current credences. Here is an advantage of NRP over RP. When you get to know that you will undergo a course of experience that could change your credence function from P' to Q, NRP says that your credence function should be the same as one that is obtained from your current function P and the Bayes factors that represents the experience in question. Moreover, I should emphasize here that, like RP, NRP also says how our future credence function is reflected in our current credence function. That is, NRP also requires that if we know what our future credence function is, then our current credence function should be the same as the future one. This can be shown easily (see app. B). From the above considerations, we can say that NRP is a generalization of RP. This means that NRP says how our future experience as well as our future credence function should be reflected in our current function, while RP says nothing but how our future credence function should be reflected.

Now, let's consider the relationship between NRP and conditionalization. First of all, it can be shown that NRP is preserved under SC. That is to say, we can show that agents who satisfied NRP before updating by SC on E also satisfy NRP after the updating (see app. C). Similarly, we can also prove that NRP is preserved under JC. Suppose that a JC updater who satisfies NRP at time t changes directly her credence in E from $P_t(E)$ to $P_{t+1}(E)$. Then, does the agent satisfy NRP at time $t + 1$? For notational simplicity, let C refer to

https://doi.org/10.1086/668001 Published online by Cambridge University Press

C_{t+b}^{t+a}. (It is assumed here that $0 < b < a$ and $P_t(C) > 0$.) Since the agent satisfies NRP at time t, we have that

$$\text{For any } X,\ P_t(X|C) = \frac{\sum_i \beta_i P_t(X \& \omega_i)}{\sum_i \beta_i P_t(\omega_i)}. \tag{13}$$

Now, let's update the conditional credence in X given C using JC on $\{E, \sim E\}$. By a bit of mathematics, we have that[10]

$$\text{For any } X,\ P_{t+1}(X|C) = \frac{\alpha P_t(X \& E|C) + P_t(X \& \sim E|C)}{\alpha P_t(E|C) + P_t(\sim E|C)}. \tag{14}$$

Here, α is an abbreviation for $[P_{t+1}(E)/P_t(E)]/[P_{t+1}(\sim E)/P_t(\sim E)]$. Appendix D shows that it follows from equations (13) and (14) that

$$\text{For any } X,\ P_{t+1}(X|C) = \frac{\sum_i \beta_i P_{t+1}(X \& \omega_i)}{\sum_i \beta_i P_{t+1}(\omega_i)}. \tag{15}$$

Finally, we can conclude that the agent also satisfies NRP after updating by JC on $\{E, \sim E\}$. That is, an agent who satisfies NRP before updating by JC also satisfies NRP after the updating.[11] As our discussion has shown, NRP is a generalization of RP and is preserved under conditionalization—SC and JC.

5. Concluding Remarks. In this article, I have discussed a prima facie conflict between the reflection principle and another constraint—conditionalization. In particular, I have shown that RP conflicts with JC and that this conflict is due to an ambiguity of the standard formulation of the principle (i.e., RP). So, I suggested a new formulation of the principle so as to circumvent the conflict in question.

Discussions in the literature about the relationship between conditionalization and the reflection principle have been limited to the relationship between SC and the reflection principle (see, e.g., van Fraassen 1984, 1995;

10. It is assumed that $P_t(C) > 0$. Note that for any proposition X, if an agent's old credence in X is greater than 0 and is updated by a JC, then her new credence in X is also greater than 0. Thus, $P_{t+1}(C) > 0$. So, $P_{t+1}(X|C)$ is well defined for any X. According to the probability calculus, thus, it holds that $P_{t+1}(X|C) = P_{t+1}(X \& C)/P_{t+1}(C)$ for any X. From JC and the probability calculus, then, it follows that for any X $P_{t+1}(X|C) = P_{t+1}(X \& C)/P_{t+1}(C) = [P_{t+1}(E)P_t(X \& C|E) + P_{t+1}(\sim E)P_t(X \& C|\sim E)]/[P_{t+1}(E)P_t(C|E) + P_{t+1}(\sim E)P_t(C|\sim E)]$. Finally, if we transform the conditional probabilities on the right-hand side by means of Bayes's theorem, we can obtain eq. (14).

11. This result can be generalized to cases in which an agent's credences in the members of a partition $\{E_1, \ldots, E_n\}$ are directly affected by a course of experience and the credence function is updated by JC on the partition. This is just a matter of the complexity of mathematics.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

Maher 1993; Howson and Urbach 2006; Weisberg 2007). And the relationship between JC and the principle has attracted less attention. As we discussed in this article, however, the standard formulation of the principle conflicts with JC, and so we need to revise the formulation. For this reason, this article may call philosophers' attention to this issue. And I want to emphasize that my proposal (i.e., NRP) is a revision of RP using Bayes factors. Some philosophers like Wagner (2000) and Jeffrey (2002) have attempted to resolve some critical problems of Bayesianism by means of Bayes factors—for example, the old evidence problem. Interestingly, I have shown that Bayes factors are one of the keys to resolving the conflict between RP and JC. For this reason, my proposal may be treated as an exemplar that shows the theoretical importance of Bayes factors.

Appendix A

Let's show that the following three equations imply that $P_0(E_1) = P_1(E_1)$:

$$P_0(R_2|E_1) = P_1(R_2|E_1); \tag{A1}$$

$$P_0(R_2|\sim E_1) = P_1(R_2|\sim E_1); \tag{A2}$$

$$P_0(E_1|R_2) = P_1(E_1|R_2). \tag{A3}$$

From Bayes's theorem, it follows that

$$\begin{aligned} P_0(E_1|R_2) &= \frac{P_0(E_1)P_0(R_2|E_1)}{P_0(E_1)P_0(R_2|E_1) + P_0(\sim E_1)P_0(R_2|\sim E_1)}, \text{ and} \\ P_1(E_1|R_2) &= \frac{P_1(E_1)P_1(R_2|E_1)}{P_1(E_1)P_1(R_2|E_1) + P_1(\sim E_1)P_1(R_2|\sim E_1)} \end{aligned} \tag{A4}$$

Let k and l be numbers such that $P_0(R_2|E_1) = P_1(R_2|E_1) = k$ and $P_0(R_2|\sim E_1) = P_1(R_2|\sim E_1) = l$. Then, equations (A1)–(A4) imply that

$$\frac{P_0(E_1)k}{P_0(E_1)k + P_0(\sim E_1)l} = \frac{P_1(E_1)k}{P_1(E_1)k + P_1(\sim E_1)l}.$$

If $k = 0$, then $P_0(E_1|R_2) = P_1(E_1|R_2) = 0$, and so $P_2(E_1) = 0$. Note that it has been assumed that $P_2(E_1)$ is in (0, 1). Thus, $k \neq 0$. Similarly, it is implied that $l \neq 0$. Thus, it follows from the last equation that $P_0(E_1) = P_1(E_1)$.

https://doi.org/10.1086/668001 Published online by Cambridge University Press

Appendix B

Suppose that $P_{t+b} = P_t$. Then, $\beta_i = [P_{t+a}(\omega_i)/P_t(\omega_i)]/[P_{t+a}(\omega_1)/P_t(\omega_1)]$. From NRP, then, it follows that

$$P_t(X|C_t^{t+a}) = \frac{\sum_i [P_{t+a}(\omega_i)/P_t(\omega_i)]/[P_{t+a}(\omega_1)/P_t(\omega_1)]P_t(X\&\omega_i)}{\sum_i [P_{t+a}(\omega_i)/P_t(\omega_i)]/[P_{t+a}(\omega_1)/P_t(\omega_1)]P_t(\omega_i)}.$$

And this equation can be transformed as follows:

$$P_t(X|C_t^{t+a}) = \frac{\sum_i [P_{t+a}(\omega_i)/P_t(\omega_i)]P_t(X\&\omega_i)}{\sum_i [P_{t+a}(\omega_i)/P_t(\omega_i)]P_t(\omega_i)}.$$

The denominator of the right-hand side is equal to 1 since $\sum_i P_{t+a}(\omega_i) = 1$. And the numerator of the right-hand side is transformed as follows: $\sum_i P_{t+a}(\omega_i)P_t(X|\omega_i)$. Note here that ω_i is a disjunct of X or not. And $P_t(X|\omega_i) = 1$ if ω_i is a disjunct of X; $P_t(X|\omega_i) = 0$, otherwise. And this logical relation is not changed at time $t + a$. Thus, $P_t(X|\omega_i) = P_{t+a}(X|\omega_i)$. Therefore, we have that $\sum_i P_{t+a}(\omega_i)P_t(X|\omega_i) = \sum_i P_{t+a}(\omega_i)P_{t+a}(X|\omega_i) = P_{t+a}(X)$. Finally, we have that $P_t(X|C_t^{t+a}) = P_{t+a}(X)$. Now, we can conclude that when $P_{t+b} = P_t$, NRP is equivalent to equation (9). As was explained, the value of $P_t(X|C_t^{t+a})$ is equal to the value of $P_t(X|R_{t+a})$. Thus, it follows that $P_t(X|R_{t+a}) = P_{t+a}(X)$. For this reason, NRP is a generalization of RP.

Appendix C

Suppose that $0 \le b < a$ and $P_t(C_{t+b}^{t+a}\&Y) > 0$. From the probability calculus and NRP, it follows that

$$\begin{aligned} P_t(X|C_{t+b}^{t+a}\&Y) &= \frac{P_t(X\&Y|C_{t+b}^{t+a})}{P_t(Y|C_{t+b}^{t+a})} \\ &= \frac{\sum_i \beta_i P_t(X\&Y\&\omega_i)/\sum_i \beta_i P_t(\omega_i)}{\sum_i \beta_i P_t(Y\&\omega_i)/\sum_i \beta_i P_t(\omega_i)} \\ &= \frac{\sum_i \beta_i P_t(X\&\omega_i|Y)}{\sum_i \beta_i P_t(\omega_i|Y)}. \end{aligned}$$

Thus, we can say NRP implies that

https://doi.org/10.1086/668001 Published online by Cambridge University Press

NRP*: For any propositions X and Y and any real number $0 \leq b < a$,

$$P_t(X|C^{t+a}_{t+b}\&Y) = \frac{\sum_i \beta_i P_t(X\&\omega_i|Y)}{\sum_i \beta_i P_t(\omega_i|Y)}, \text{ when } P_t(C^{t+a}_{t+b}\&Y) > 0.$$

Now let's prove that NRP is preserved under SC. Suppose that an agent satisfies NRP at time t. Suppose also that she learns that E is true at time $t + 1$, and so she updates her credence using SC on E. As was mentioned, NRP implies NRP*. Thus, it follows from NRP that, for any X,

$$P_t(X|C^{t+a}_{t+b}\&E) = \frac{\sum_i \beta_i P_t(X\&\omega_i|E)}{\sum_i \beta_i P_t(\omega_i|E)}. \tag{C1}$$

Here it is assumed that $0 < b < a$ and $P_t(C^{t+a}_{t+b}\&E) > 0$. Note that it follows from SC that, for any X,

$$\begin{aligned} P_t(X|C^{t+a}_{t+b}\&E) &= P_{t+1}(X|C^{t+a}_{t+b}); \\ P_t(X\&\omega_i|E) &= P_{t+1}(X\&\omega_i); \\ P_t(\omega_i|E) &= P_{t+1}(\omega_i). \end{aligned} \tag{C2}$$

From equations (C1) and (C2), then, it follows that

$$P_{t+1}(X|C^{t+a}_{t+b}) = \frac{\sum_i \beta_i P_{t+1}(X\&\omega_i)}{\sum_i \beta_i P_{t+1}(\omega_i)}.$$

Thus, the agent satisfies NRP at time $t + 1$.

Appendix D

From equation (13), we have the following equations:

$$\begin{aligned} P_t(X\&E|C) &= \frac{\sum_i \beta_i P_t(X\&E\&\omega_i)}{\sum_i \beta_i P_t(\omega_i)}; \\ P_t(X\&\sim E|C) &= \frac{\sum_i \beta_i P_t(X\&\sim E\&\omega_i)}{\sum_i \beta_i P_t(\omega_i)}; \\ P_t(E|C) &= \frac{\sum_i \beta_i P_t(E\&\omega_i)}{\sum_i \beta_i P_t(\omega_i)}; \\ P_t(\sim E|C) &= \frac{\sum_i \beta_i P_t(\sim E\&\omega_i)}{\sum_i \beta_i P_t(\omega_i)}. \end{aligned} \tag{D1}$$

https://doi.org/10.1086/668001 Published online by Cambridge University Press

Then, the above equations and (14) imply that

$$P_{t+1}(X|C) = \frac{\alpha\sum_i\beta_i P_t(X\&E\&\omega_i) + \sum_i\beta_i P_t(X\&\sim E\&\omega_i)}{\alpha\sum_i\beta_i P_t(E\&\omega_i) + \sum_i\beta_i P_t(\sim E\&\omega_i)}. \tag{D2}$$

Note that $\alpha = [P_{t+1}(E)/P_t(E)]/[P_{t+1}(\sim E)/P_t(\sim E)]$. Then, equation (D2) is transformed as follows:

$$\begin{aligned} & P_{t+1}(X|C) \\ &= \frac{\sum_i\{[P_{t+1}(E)]/[P_t(E)]\}\beta_i P_t(X\&E\&\omega_i) + \sum_i\{[P_{t+1}(\sim E)]/[P_t(\sim E)]\}\beta_i P_t(X\&\sim E\&\omega_i)}{\sum_i\{[P_{t+1}(E)]/[P_t(E)]\}\beta_i P_t(E\&\omega_i) + \sum_i\{[P_{t+1}(\sim E)]/[P_t(\sim E)]\}\beta_i P_t(\sim E\&\omega_i)} \\ &= \frac{\sum_i P_{t+1}(E)\beta_i P_t(X\&\omega_i|E) + \sum_i P_{t+1}(\sim E)\beta_i P_t(X\&\omega_i|\sim E)}{\sum_i P_{t+1}(E)\beta_i P_t(\omega_i|E) + \sum_i P_{t+1}(\sim E)\beta_i P_t(\omega_i|\sim E)} \\ &= \frac{\sum_i\beta_i[P_{t+1}(E)P_t(X\&\omega_i|E) + P_{t+1}(\sim E)P_t(X\&\omega_i|\sim E)]}{\sum_i\beta_i[P_{t+1}(E)P_t(\omega_i|E) + P_{t+1}(\sim E)P_t(\omega_i|\sim E)]}. \end{aligned} \tag{D3}$$

Meanwhile, JC says that

$$\begin{aligned} P_{t+1}(X\&\omega_i) &= P_{t+1}(E)P_t(X\&\omega_i|E) + P_{t+1}(\sim E)P_t(X\&\omega_i|\sim E); \\ P_{t+1}(\omega_i) &= P_{t+1}(E)P_t(\omega_i|E) + P_{t+1}(\sim E)P_t(\omega_i|\sim E). \end{aligned}$$

Thus, equation (D3) is transformed as follows:

$$P_{t+1}(X|C) = \frac{\sum_i\beta_i P_{t+1}(X\&\omega_i)}{\sum_i\beta_i P_{t+1}(\omega_i)}.$$

Finally, we have equation (15).

REFERENCES

Field, Hartry. 1978. "A Note on Jeffrey Conditionalization." *Philosophy of Science* 45:361–67.

Howson, Colin, and Peter Urbach. 2006. *Scientific Reasoning: The Bayesian Approach*. 2nd ed. Chicago: Open Court.

Jeffrey, Richard. 2002. *Subjective Probability: The Real Thing*. Cambridge: Cambridge University Press.

Maher, Patrick. 1993. *Betting on Theories*. Cambridge: Cambridge University Press.

van Fraassen, Bas. 1984. "Belief and the Will." *Journal of Philosophy* 8 (1): 235–56.

———. 1995. "Belief and the Problem of Ulysses and the Sirens." *Philosophical Studies* 77:7–37.

Wagner, Carl. 2000. "Old Evidence and New Explanation III." *Philosophy of Science* 68 (Proceedings): S165–S175.

———. 2002. "Probability Kinematics and Commutativity." *Philosophy of Science* 69:266–78.

———. 2003. "Commuting Probability Revisions: The Uniformity Rule." *Erkenntnis* 59:349–64.

Weisberg, Jonathan. 2007. "Conditionalization, Reflection, and Self-Knowledge." *Philosophical Studies* 135:179–97.

https://doi.org/10.1086/668001 Published online by Cambridge University Press