What Conditional Probability Must (Almost) Be

Kenny Easwaran

Group in Logic and Methodology of Science
University of California, Berkeley

May 25, 2005
Hájek’s Argument

The Borel Paradox
  
  My Notation
  Two Intuitions on Conditional Probabilities
  Vindication of the Second Intuition
  The Paradox (and a Resolution)

Results
  A Generalization of the Resolution
  A Three-Place Function
  An Argument for Relativization
Kolmogorov took unconditional probabilities as basic and said $P(A|B) = \frac{P(A \& B)}{P(B)}$

Starting with Popper some have said that conditional probability should be basic and we should let $P(A) = P(A|T)$ where $T$ is some tautology.

Hájek points out that conditional probability is not merely a technical term, but a pre-theoretic one, so either of these is an analysis of conditional probability, and not actually a definition.

Hájek claims that we should side with the latter analysis.
I argue that it is *impossible* that there be conditional probabilities for every pair of events, whether unconditional or conditional probability is taken as basic.

This will depend on a reflection principle, stating that if $B$ is a union of some events $E_\alpha$, then $P(A|B) \geq \min\{P(A|E_\alpha)\}$.

However, there is a three-place function that does satisfy all the traditional axioms as well as this reflection principle.

I prove that any such three-place function must be almost equal to one that Kolmogorov describes later in his book.
Hájek’s Argument

The Borel Paradox
  My Notation
  Two Intuitions on Conditional Probabilities
  Vindication of the Second Intuition
  The Paradox (and a Resolution)

Results
  A Generalization of the Resolution
  A Three-Place Function
  An Argument for Relativization
Four Horn Theorem - any space with uncountably many disjoint events must assign uncountably many of them either zero, infinitesimal, vague, or no probability
Four Horn Theorem - any space with uncountably many disjoint events must assign uncountably many of them either zero, infinitesimal, vague, or no probability

If $B$ is such an event then $P(A|B)$ is undefined on Kolmogorov’s ratio analysis
Four Horn Theorem - any space with uncountably many disjoint events must assign uncountably many of them either zero, infinitesimal, vague, or no probability

If $B$ is such an event then $P(A|B)$ is undefined on Kolmogorov’s ratio analysis

However, in some such cases, we have a clear intuition as to what $P(A|B)$ should be
Four Horn Theorem - any space with uncountably many disjoint events must assign uncountably many of them either zero, infinitesimal, vague, or no probability

If $B$ is such an event then $P(A|B)$ is undefined on Kolmogorov’s ratio analysis

However, in some such cases, we have a clear intuition as to what $P(A|B)$ should be

Therefore, Kolmogorov’s ratio analysis is incorrect
“The examples of vague and undefined probabilities suggest that the problem with the ratio analysis is not that it is a ratio analysis, as opposed to some other function of unconditional probabilities. The problem lies in the very attempt to analyze conditional probabilities in terms of unconditional probabilities at all. It seems that any other putative analysis that treated unconditional probability as more basic than conditional probability would meet a similar fate - as Kolmogorov’s elaboration (RV) did.”
But I will argue that Kolmogorov’s elaboration is actually *better* than basic conditional probabilities. The ratio analysis undergenerates, but basic conditional probability overgenerates.

In addition, I will show that *any* attempted analysis of conditional probability must almost equal Kolmogorov’s elaboration.
Hájek’s Argument

The Borel Paradox

My Notation
Two Intuitions on Conditional Probabilities
Vindication of the Second Intuition
The Paradox (and a Resolution)

Results
A Generalization of the Resolution
A Three-Place Function
An Argument for Relativization
“Suppose that we have a uniform probability measure over the Earth’s surface (imagine it to be a perfect sphere). What is the probability that a randomly chosen point lies in the western hemisphere (W), given that it lies on the equator (E)? 1/2, surely. But the probability that the point lies on the equator is 0, since the equator has no area. . . . We could have asked a similar question for each of the uncountably many lines of latitude, with the same answer in each case.”
Consider a sphere of surface area 1, with center $O$ and let $X$ be a point chosen uniformly at random from the surface.

For any set $E$, the probability of $E$ is the area of $E$.

I will (merely for ease of notation) occasionally identify events with the region in which they are true, and occasionally with the set of possible worlds in which they are true. I don’t mean to presuppose that events are in fact any of these things.
Consider a sphere of surface area 1, with center $O$ and let $X$ be a point chosen uniformly at random from the surface.

For any set $E$, the probability of $E$ is the area of $E$.

I will (merely for ease of notation) occasionally identify events with the region in which they are true, and occasionally with the set of possible worlds in which they are true. I don’t mean to presuppose that events are in fact any of these things.
If $A$ is a point on the sphere, and $0 \leq \theta_0 \leq \theta_1 \leq \pi$ then $S_{A,\theta_0,\theta_1}$ occurs just in case $\theta_0 \leq \angle XO A \leq \theta_1$
If $A$ is a point on the sphere, and $0 \leq \theta_0 \leq \theta_1 \leq \pi$ then $S_{A,\theta_0,\theta_1}$ occurs just in case $\theta_0 \leq \angle XOA \leq \theta_1$

$P(S_{A,\theta_0,\theta_1}) = (\cos \theta_0 - \cos \theta_1)/2$ by a simple integration
If $A$ is a point on the sphere, and $0 \leq \theta_0 \leq \theta_1 \leq \pi$ then $S_{A,\theta_0,\theta_1}$ occurs just in case $\theta_0 \leq \angle XOA \leq \theta_1$

$P(S_{A,\theta_0,\theta_1}) = (\cos \theta_0 - \cos \theta_1)/2$ by a simple integration

$C_A$ occurs just in case $\angle XOA = \pi/2$, so that $X$ lies on the equator if $A$ is a pole
If $A$ is a point on the sphere, and $0 \leq \theta_0 \leq \theta_1 \leq \pi$ then $S_{A,\theta_0,\theta_1}$ occurs just in case $\theta_0 \leq \angle XOA \leq \theta_1$

$P(S_{A,\theta_0,\theta_1}) = (\cos \theta_0 - \cos \theta_1)/2$ by a simple integration

$C_A$ occurs just in case $\angle XOA = \pi/2$, so that $X$ lies on the equator if $A$ is a pole

$P(C_A) = 0$ because a line has zero area
First Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
First Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
- $C_B$ occurs just in case $X$ is on the Greenwich meridian
- By consideration of lengths, it seems that $P(A|C_B) = 1/6$
First Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
- $C_B$ occurs just in case $X$ is on the Greenwich meridian
- By consideration of lengths, it seems that $P(A|C_B) = 1/6$
- In general, it seems that $P(S_{N,\theta_0,\theta_1}|C_B) = \frac{\theta_1 - \theta_0}{\pi}$ whenever $\angle AOB = \pi/2$
Second Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
Second Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N, 0, \pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
- $C_B$ occurs just in case $X$ is on the Greenwich meridian
- By symmetry, it shouldn’t matter which longitude $C_B$ picks out, so $A$ should be independent of $C_B$
Second Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)
- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north
- $C_B$ occurs just in case $X$ is on the Greenwich meridian
- By symmetry, it shouldn’t matter which longitude $C_B$ picks out, so $A$ should be independent of $C_B$
- That is, $P(A|C_B) = P(A) = \frac{2-\sqrt{3}}{4}$
Second Intuition

- Let $N$ be the north pole and $B$ be at 90 degrees west longitude, 0 degrees latitude (on the equator)

- $A = S_{N,0,\pi/6}$ occurs just in case $X$ has latitude above 60 degrees north

- $C_B$ occurs just in case $X$ is on the Greenwich meridian

- By symmetry, it shouldn’t matter which longitude $C_B$ picks out, so $A$ should be independent of $C_B$

- That is, $P(A|C_B) = P(A) = \frac{2-\sqrt{3}}{4}$

- In general, it seems that $P(S_{N,\theta_0,\theta_1}|C_B) = \frac{\cos \theta_0 - \cos \theta_1}{2}$ whenever $\angle AOB = \pi/2$
Combining these two general intuitions tells us that

\[ \frac{\theta_1 - \theta_0}{\pi} = \frac{\cos \theta_0 - \cos \theta_1}{2}, \]

that is,

\[ \frac{\cos \theta_1 - \cos \theta_0}{\theta_1 - \theta_0} = -\frac{2}{\pi}. \]

The solutions to this equation are just pairs of θ-coordinates of intersections of a line of slope \(-2/\pi\) with the graph of \(y = \cos \theta\), but each value between 0 and \(\pi\) occurs in at most three such pairs, so this equation is almost never satisfied.

Therefore, almost always, these intuitions can’t both be correct, which is the paradox.
Let $E$ be the partition of the sphere into the lines of longitude

Let $A$ be any set $S_{N,\theta_0,\theta_1}$ as above, where $N$ is the north pole

Assume that $P(A|E_\alpha)$ is defined for every line of longitude $E_\alpha$

I will show that for almost every line $E_\alpha$, we must have

$$P(A|E_\alpha) = P(A) = \frac{\cos \theta_0 - \cos \theta_1}{2},$$

as per the second intuition.
For any point \( w \) on the surface of the sphere, define
\[ h(w) = P(A|E_\alpha) - P(A), \]
where \( E_\alpha \) is the unique line of longitude containing \( w \). (If \( w \) is a pole, then let \( h(w) = 0 \).)

Let \( B \) be the union of the lines \( E_\alpha \) such that \( P(A|E_\alpha) \neq P(A) \).
For any point $w$ on the surface of the sphere, define $h(w) = P(A|E_\alpha) - P(A)$, where $E_\alpha$ is the unique line of longitude containing $w$. (If $w$ is a pole, then let $h(w) = 0$.)

Let $B$ be the union of the lines $E_\alpha$ such that $P(A|E_\alpha) \neq P(A)$.

Then $P(B) = P(h(w) \neq 0)$.

Thus, if we can show that $P(h(w) = 0) = 1$, then we are done.
If $P(h(w) \neq 0) > 0$, then either $P(h(w) > 0) > 0$ or $P(h(w) < 0) > 0$

Without loss of generality, assume the former

Assuming countable additivity, we see that there is some positive $\epsilon$ such that $P(h(w) > \epsilon)$ is positive

Let $B_\epsilon$ be the event that $h(w) > \epsilon$. Since $h = P(A|E_\alpha) - P(A)$, and $E_\alpha$ are the lines of longitude, this event is a union of lines of longitude.
Now consider $P(A|B_\epsilon)$

By definition, we know that for any $E_\alpha$ contained in $B_\epsilon$, we have $P(A|E_\alpha) > P(A) + \epsilon$, since $B_\epsilon$ is the region where $h > \epsilon$

The reflection principle stated earlier would thus suggest that $P(A|B_\epsilon) \geq P(A) + \epsilon$, since this is a disjoint union
A Reflection Principle

> If $B_\epsilon$ is the disjoint union of the $E_\alpha$, then I will argue that $P(A|B_\epsilon) \geq \min\{P(A|E_\alpha)\}$
A Reflection Principle

- If $B_\varepsilon$ is the disjoint union of the $E_\alpha$, then I will argue that $P(A|B_\varepsilon) \geq \min\{P(A|E_\alpha)\}$

- If we learn that the point $X$ is in $B_\varepsilon$ and nothing more, then our current probability for $A$ should be $P(A|B_\varepsilon)$
A Reflection Principle

- If $B_\varepsilon$ is the disjoint union of the $E_\alpha$, then I will argue that
  \[ P(A|B_\varepsilon) \geq \min\{P(A|E_\alpha)\} \]
- If we learn that the point $X$ is in $B_\varepsilon$ and nothing more, then our current probability for $A$ should be $P(A|B_\varepsilon)$
- If we learn further that the line of longitude $X$ is on is $E_\alpha$, then our probability for $A$ should become $P(A|E_\alpha)$
- But if $P(A|B_\varepsilon) = \min\{P(A|E_\alpha)\} - \delta$, then this results in an increase of at least $\delta$ in our probability for $A$ no matter which $E_\alpha$ we discover it to be on.
- Since we can make this increase regardless of which line of longitude we learn the point to be on, it seems that we should be able to do it already.
A Reflection Principle

- If $B_\epsilon$ is the disjoint union of the $E_\alpha$, then I will argue that $P(A|B_\epsilon) \geq \min\{P(A|E_\alpha)\}$
- If we learn that the point $X$ is in $B_\epsilon$ and nothing more, then our current probability for $A$ should be $P(A|B_\epsilon)$
- If we learn further that the line of longitude $X$ is on is $E_\alpha$, then our probability for $A$ should become $P(A|E_\alpha)$
- But if $P(A|B_\epsilon) = \min\{P(A|E_\alpha)\} - \delta$, then this results in an increase of at least $\delta$ in our probability for $A$ no matter which $E_\alpha$ we discover it to be on
A Reflection Principle

- If $B_\epsilon$ is the disjoint union of the $E_\alpha$, then I will argue that $P(A|B_\epsilon) \geq \min\{P(A|E_\alpha)\}$
- If we learn that the point $X$ is in $B_\epsilon$ and nothing more, then our current probability for $A$ should be $P(A|B_\epsilon)$
- If we learn further that the line of longitude $X$ is on is $E_\alpha$, then our probability for $A$ should become $P(A|E_\alpha)$
- But if $P(A|B_\epsilon) = \min\{P(A|E_\alpha)\} - \delta$, then this results in an increase of at least $\delta$ in our probability for $A$ no matter which $E_\alpha$ we discover it to be on
- Since we can make this increase regardless of which line of longitude we learn the point to be on, it seems that we should be able to do it already
Thus, by this reflection principle, if $B_\epsilon$ is the region where $h > \epsilon$, then $P(A \mid B_\epsilon) \geq P(A) + \epsilon$. However, it is not hard to verify that if $A$ is composed entirely of lines of latitude and $B$ is composed entirely of lines of longitude, then $P(A \land B) = P(A) P(B)$. Thus, a contradiction arises from assuming that $P(h > \epsilon) > 0$, so $P(A \mid E_\alpha)$ is almost always almost equal to $P(A)$ (the second "almost" can be dropped if we assume countable additivity). This is what the second intuition told us by a simple symmetry argument.
Thus, by this reflection principle, if $B_\epsilon$ is the region where $h > \epsilon$, then $P(A|B_\epsilon) \geq P(A) + \epsilon$.

Since $B_\epsilon$ has positive probability, we also see that $P(A|B_\epsilon) = \frac{P(A \& B_\epsilon)}{P(B_\epsilon)}$. 
Thus, by this reflection principle, if $B_\epsilon$ is the region where $h > \epsilon$, then $P(A|B_\epsilon) \geq P(A) + \epsilon$.

Since $B_\epsilon$ has positive probability, we also see that

$$P(A|B_\epsilon) = \frac{P(A \& B_\epsilon)}{P(B_\epsilon)}$$

Multiplying through, we see that

$$P(A \& B_\epsilon) \geq P(A)P(B_\epsilon) + \epsilon P(B_\epsilon) > P(A)P(B_\epsilon)$$
Thus, by this reflection principle, if $B_{\epsilon}$ is the region where $h > \epsilon$, then $P(A|B_{\epsilon}) \geq P(A) + \epsilon$

Since $B_{\epsilon}$ has positive probability, we also see that $P(A|B_{\epsilon}) = \frac{P(A \& B_{\epsilon})}{P(B_{\epsilon})}$

Multiplying through, we see that $P(A \& B_{\epsilon}) \geq P(A)P(B_{\epsilon}) + \epsilon P(B_{\epsilon}) > P(A)P(B_{\epsilon})$

However, it is not hard to verify that if $A$ is composed entirely of lines of latitude and $B$ is composed entirely of lines of longitude, then $P(A \& B) = P(A)P(B)$

Thus, a contradiction arises from assuming that $P(h > \epsilon) > 0$, so $P(A|E_{\alpha})$ is almost always almost equal to $P(A)$ (the second "almost" can be dropped if we assume countable additivity)
Thus, by this reflection principle, if $B_\epsilon$ is the region where $h > \epsilon$, then $P(A|B_\epsilon) \geq P(A) + \epsilon$

Since $B_\epsilon$ has positive probability, we also see that $P(A|B_\epsilon) = \frac{P(A \& B_\epsilon)}{P(B_\epsilon)}$

Multiplying through, we see that $P(A \& B_\epsilon) \geq P(A)P(B_\epsilon) + \epsilon P(B_\epsilon) > P(A)P(B_\epsilon)$

However, it is not hard to verify that if $A$ is composed entirely of lines of latitude and $B$ is composed entirely of lines of longitude, then $P(A \& B) = P(A)P(B)$

Thus, a contradiction arises from assuming that $P(h > \epsilon) > 0$, so $P(A|E_\alpha)$ is almost always almost equal to $P(A)$ (the second “almost” can be dropped if we assume countable additivity)

This is what the second intuition told us by a simple symmetry argument
The first intuition has the property that \( P(\cdot | E_\alpha) \) gives uniform random distribution over the line \( E_\alpha \).

The second intuition suggests that for a uniform distribution over the sphere, conditioning on some great circle yields a non-uniform distribution on the circle.

But is there a reason to believe conditional probabilities should depend on lengths when the unconditional ones depend on areas?
The second intuition satisfies the reflection property that if $B$ is the union of disjoint events $E_\alpha$ and $x < P(A|E_\alpha) < y$ for all $\alpha$, then $x \leq P(A|B) \leq y$.

The first intuition says $P(S_{A,0,\pi/6}|C_B) = 1/6$ for any line of longitude $C_B$, while $P(S_{A,0,\pi/6}) = \frac{2-\sqrt{3}}{2}$ is strictly less.

Thus, an agent violating the second intuition would set herself up for an infinitary Dutch book by violating the reflection principle.
The second intuition satisfies the reflection property that if $B$ is the union of disjoint events $E_\alpha$ and $x < P(A|E_\alpha) < y$ for all $\alpha$, then $x \leq P(A|B) \leq y$

The first intuition says $P(S_{A,0,\pi/6}|C_B) = 1/6$ for any line of longitude $C_B$, while $P(S_{A,0,\pi/6}) = \frac{2-\sqrt{3}}{2}$ is strictly less

Thus, an agent violating the second intuition would set herself up for an infinitary Dutch book by violating the reflection principle:

She gets $1$ if $S_{A,0,\pi/6}$ occurs, and I get $\frac{2}{2-\sqrt{3}}$ if it doesn’t

For each $E_\alpha$, I get $1$ if both $E_\alpha$ and $S_{A,0,\pi/6}$ occur, and she gets $6$ if $E_\alpha$ occurs and $S_{A,0,\pi/6}$ doesn’t
The second intuition satisfies the reflection property that if \( B \) is the union of disjoint events \( E_\alpha \) and \( x < P(A|E_\alpha) < y \) for all \( \alpha \), then \( x \leq P(A|B) \leq y \).

The first intuition says \( P(S_A,0,\pi/6|C_B) = 1/6 \) for any line of longitude \( C_B \), while \( P(S_A,0,\pi/6) = \frac{2 - \sqrt{3}}{2} \) is strictly less.

Thus, an agent violating the second intuition would set herself up for an infinitary Dutch book by violating the reflection principle:

- She gets $1 if \( S_{A,0,\pi/6} \) occurs, and I get $\frac{2}{2 - \sqrt{3}}$ if it doesn’t.
- For each \( E_\alpha \), I get $1 if both \( E_\alpha \) and \( S_{A,0,\pi/6} \) occur, and she gets $6 if \( E_\alpha \) occurs and \( S_{A,0,\pi/6} \) doesn’t.
- If \( S_{A,0,\pi/6} \) occurs, then this means we break even, and if it doesn’t, then I get some positive amount (since \( \frac{2}{2 - \sqrt{3}} > 6 \)), so by scaling the first bet down slightly, I can win either way.
My Resolution of the Paradox

- The first intuition is supported by some mental associations between length and area in a uniform distribution.
- In a non-uniform distribution, such an intuition disappears entirely.
- But the reflection principle and its (possibly infinitary) Dutch book remain.
- Thus I claim the second intuition is the right one, since it can be generalized.
Hájek’s Argument

The Borel Paradox

My Notation
Two Intuitions on Conditional Probabilities
Vindication of the Second Intuition
The Paradox (and a Resolution)

Results

A Generalization of the Resolution
A Three-Place Function
An Argument for Relativization
In the case of the Borel paradox, I argued that for almost all lines of longitude $E_{\alpha}$, with $A$ some region symmetric around the pole, we had to have $P(A|E_{\alpha}) = P(A)$ (that is, if $B_\varepsilon$ is the union of the $E_{\alpha}$ on which this equation is off by $\varepsilon$, then $P(B_\varepsilon) = 0$).

This was proved by showing that for any $B$ composed of a union of the $E_{\alpha}$, we must have $P(A&B) = \int_B f(w) \, dw$, where $f(w) = P(A|E_{\alpha})$ for the unique $E_{\alpha}$ containing $w$.

If we use the Lebesgue definition of the integral instead of the Riemann definition, then we can generalize this result to other probability spaces.
Let $E$ be some partition of an arbitrary probability space into events $E_\alpha$

Define $f(w) = P(A|E_\alpha)$, where $E_\alpha$ is the unique element of the partition containing $w$, as above
Let $E$ be some partition of an arbitrary probability space into events $E_\alpha$.

Define $f(w) = P(A|E_\alpha)$, where $E_\alpha$ is the unique element of the partition containing $w$, as above.

If $B$ is the union of finitely many $E_\alpha$, then we see that

$$P(A\&B) = \sum P(A|E_\alpha)P(E_\alpha) = \sum f(w)P(E_\alpha)$$
Let $E$ be some partition of an arbitrary probability space into events $E_\alpha$.

Define $f(w) = P(A|E_\alpha)$, where $E_\alpha$ is the unique element of the partition containing $w$, as above.

If $B$ is the union of finitely many $E_\alpha$, then we see that

$$P(A \& B) = \sum P(A|E_\alpha)P(E_\alpha) = \sum f(w)P(E_\alpha)$$

But this just means $P(A \& B) = \int_B f(w)dw$ as claimed before.
Let $E$ be some partition of an arbitrary probability space into events $E_\alpha$

Define $f(w) = P(A|E_\alpha)$, where $E_\alpha$ is the unique element of the partition containing $w$, as above

If $B$ is the union of finitely many $E_\alpha$, then we see that

$$P(A\&B) = \sum P(A|E_\alpha)P(E_\alpha) = \sum f(w)P(E_\alpha)$$

But this just means $P(A\&B) = \int_B f(w)dw$ as claimed before

In fact, one can show (even without assuming countable additivity) that this integral equation holds when $B$ is any union of the $E_\alpha$, though it is necessary to appeal to a generalization of the reflection principle appealed to in the discussion of the Borel paradox

(The proof occurs in the first appendix of my written version on the FEW website)
Now let $g_A$ be any function constant on each $E_\alpha$ such that $P(A \& B) = \int_B g(w)dw$ for all $B$ that are unions of some $E_\alpha$.

Let $f(w) = P(A|E_\alpha)$ whenever $w \in E_\alpha$, as above.
Now let $g_A$ be any function constant on each $E_\alpha$ such that $P(A\&B) = \int_B g(w)dw$ for all $B$ that are unions of some $E_\alpha$.

Let $f(w) = P(A|E_\alpha)$ whenever $w \in E_\alpha$, as above.

It is now possible to show (again using the reflection principle), that the $P(f(w) - g(w) > \epsilon) = 0$ for any positive $\epsilon$.

(The proof occurs in the second appendix of my written version on the FEW website.)

If we assume countable additivity, this condition translates to $P(f(w) = g(w)) = 1$.
Now let $g_A$ be any function constant on each $E_\alpha$ such that $P(A\&B) = \int_B g(w)dw$ for all $B$ that are unions of some $E_\alpha$.

Let $f(w) = P(A\mid E_\alpha)$ whenever $w \in E_\alpha$, as above.

It is now possible to show (again using the reflection principle), that the $P(f(w) - g(w) > \epsilon) = 0$ for any positive $\epsilon$.

(The proof occurs in the second appendix of my written version on the FEW website.)

If we assume countable additivity, this condition translates to $P(f(w) = g(w)) = 1$.

Thus, if such a $g_A$ exists, then it is what conditional probability must almost be (since conditional probabilities must satisfy this integral equation, we see that if no such $g_A$ exists, then the conditional probabilities can’t exist either).
Thus, if some function $g_A$ exists that satisfies the integral equation for conditional probability of $A$, and conditional probabilities exist for $A$ given each element of a partition, then these conditional probabilities must almost equal $g_A$. However, note that if $P(E_\alpha) = 0$, then $P(A|E_\alpha)$ may diverge from $g_A$ on this set $E_\alpha$. But this can only happen on relatively few of the $E_\alpha$. 
Thus, if some function $g_A$ exists that satisfies the integral equation for conditional probability of $A$, and conditional probabilities exist for $A$ given each element of a partition, then these conditional probabilities must almost equal $g_A$.

However, note that if $P(E_\alpha) = 0$, then $P(A|E_\alpha)$ may diverge from $g_A$ on this set $E_\alpha$. But this can only happen on relatively few of the $E_\alpha$. 

Kenny Easwaran
What Conditional Probability Must (Almost) Be
In the case of the Borel paradox, we considered a set $A = S_{N,\theta_0,\theta_1}$ where $N$ is the north pole and the partition of the space into $C_B$ where $B$ ranged over points on the equator.

A function $g_A$ was available that satisfied the integral equation for this partition, namely the constant function $P(A)$.

This was the value the second intuition gave us for $P(A|E)$. 
Now let \( N \) be the north pole, \( \tilde{N} \) be the south pole, \( B \) be some point on the equator, and \( C \) be some other point on the equator at 90 degrees from both \( N \) and \( B \).

Let \( A_1 = S_{N,0,\pi/6} \cup S_{\tilde{N},0,\pi/6} \) and let \( A_2 = S_{C,\pi/3,2\pi/3} \).
Now let $N$ be the north pole, $\bar{N}$ be the south pole, $B$ be some point on the equator, and $C$ be some other point on the equator at 90 degrees from both $N$ and $B$

Let $A_1 = S_{N,0,\pi/6} \cup S_{\bar{N},0,\pi/6}$ and let $A_2$ be $S_{c,\pi/3,2\pi/3}$

Then $C_B \cap A_1 = C_B \cap A_2$, and thus $P(A_1|C_B)$ should equal $P(A_2|C_B)$
Now let $N$ be the north pole, $\bar{N}$ be the south pole, $B$ be some point on the equator, and $C$ be some other point on the equator at 90 degrees from both $N$ and $B$

Let $A_1 = S_{N,0,\pi/6} \cup S_{\bar{N},0,\pi/6}$ and let $A_2 = S_{C,\pi/3,2\pi/3}$

Then $C_B \cap A_1 = C_B \cap A_2$, and thus $P(A_1|C_B)$ should equal $P(A_2|C_B)$

If this doesn’t hold, then it seems that $P(A_1 \& \neg A_2|C_B)$ is positive, even though that event is impossible.
When $C_B$ is considered as an element of the partition into lines of longitude, the earlier argument suggested

$$P(A_1|C_B) = P(A_1)$$
When $C_B$ is considered as an element of the partition into lines of longitude, the earlier argument suggested

\[ P(A_1|C_B) = P(A_1) \]

When $C_B$ is considered as an element of the partition into great circles through $C$, the earlier argument suggested

\[ P(A_2|C_B) = P(A_2) \]
When $C_B$ is considered as an element of the partition into lines of longitude, the earlier argument suggested

$$P(A_1|C_B) = P(A_1)$$

When $C_B$ is considered as an element of the partition into great circles through $C$, the earlier argument suggested

$$P(A_2|C_B) = P(A_2)$$

But $P(A_1) = 2 - \sqrt{3}$ and $P(A_2) = 1/2$

This seems to give a contradiction
Any account that takes conditional probabilities as basic says that $P(A_i|C_B)$ should take on one particular value.
Any account that takes conditional probabilities as basic says that \( P(A_i|C_B) \) should take on one particular value.

The (discredited) first intuition gives the value 1/3.

The (supported) second intuition gives two different values when considering different axes of rotational symmetry.
Any account that takes conditional probabilities as basic says that $P(A_i|C_B)$ should take on one particular value.

The (discredited) first intuition gives the value 1/3.

The (supported) second intuition gives two different values when considering different axes of rotational symmetry.

Thus, the only intuition that can be generalized gives two distinct answers for this case, and so any account that gives a unique answer must be wrong.
Any account that takes conditional probabilities as basic says that $P(A_i|C_B)$ should take on one particular value.

The (discredited) first intuition gives the value $1/3$.

The (supported) second intuition gives two different values when considering different axes of rotational symmetry.

Thus, the only intuition that can be generalized gives two distinct answers for this case, and so any account that gives a unique answer must be wrong.

In particular, conditional probability could not be basic in the sense that Hájek claims.

Instead, I suggest that we should consider conditional probability to be a *relative* function, rather than an absolute one as Hájek wants.
That is, it should be $P(A|B, E)$, where $E$ is a partition of the space of which $B$ is an element, rather than merely $P(A|B)$. If $P(B) > 0$, then the value required by the above arguments for $P(A|B, E)$ is independent of the choice of $E$ and must always equal $P(A \cap B)/P(B)$. Thus, relativization to the third place only becomes necessary when the antecedent has probability 0, justifying the ratio analysis when it is defined. Similarly, all the examples Hájek uses in his paper take on their probabilities independently of the partition $E$ of alternatives, so this account can support those intuitions as well. But it remains silent where it should, giving values only relative to the partition rather than absolutely, in the cases like those above.
That is, it should be $P(A|B, E)$, where $E$ is a partition of the space of which $B$ is an element, rather than merely $P(A|B)$.

If $P(B) > 0$, then the value required by the above arguments for $P(A|B, E)$ is independent of the choice of $E$ and must always equal $P(A&B)/P(B)$.

Thus, relativization to the third place only becomes necessary when the antecedent has probability 0, justifying the ratio analysis when it is defined.
That is, it should be $P(A|B, E)$, where $E$ is a partition of the space of which $B$ is an element, rather than merely $P(A|B)$.

If $P(B) > 0$, then the value required by the above arguments for $P(A|B, E)$ is independent of the choice of $E$ and must always equal $P(A&B)/P(B)$.

Thus, relativization to the third place only becomes necessary when the antecedent has probability 0, justifying the ratio analysis when it is defined.

Similarly, all the examples Hájek uses in his paper take on their probabilities independently of the partition $E$ of alternatives, so this account can support those intuitions as well.
That is, it should be $P(A|B, E)$, where $E$ is a partition of the space of which $B$ is an element, rather than merely $P(A|B)$.

If $P(B) > 0$, then the value required by the above arguments for $P(A|B, E)$ is independent of the choice of $E$ and must always equal $P(A&B)/P(B)$.

Thus, relativization to the third place only becomes necessary when the antecedent has probability 0, justifying the ratio analysis when it is defined.

Similarly, all the examples Hájek uses in his paper take on their probabilities independently of the partition $E$ of alternatives, so this account can support those intuitions as well.

But it remains silent where it should, giving values only relative to the partition rather than absolutely, in the cases like those above.
If conditional probabilities exist, then they must satisfy the integral equation $P(A\&B) = \int_B f(w)dw$, where $f(w)$ is $P(A|E_\alpha)$ for the unique $E_\alpha$ from the partition containing $w$.
If conditional probabilities exist, then they must satisfy the integral equation $P(A\&B) = \int_B f(w)dw$, where $f(w)$ is $P(A|E_\alpha)$ for the unique $E_\alpha$ from the partition containing $w$.

If we assume countable additivity, then such a function always exists (by the Radon-Nikodym theorem). In the Borel paradox, such a function can be shown to exist even without countable additivity.
If conditional probabilities exist, then they must satisfy the integral equation $P(A \& B) = \int_B f(w) \, dw$, where $f(w)$ is $P(A \mid E_\alpha)$ for the unique $E_\alpha$ from the partition containing $w$.

If we assume countable additivity, then such a function always exists (by the Radon-Nikodym theorem). In the Borel paradox, such a function can be shown to exist even without countable additivity.

If we think that $A_1 \cap B = A_2 \cap B$ implies that $P(A_1 \mid B) = P(A_2 \mid B)$, then the function specified must be a three-place function rather than a two-place one as it is traditionally considered.
Rényi, trying to argue that conditional probabilities should be taken as basic, said,
“in general, it makes sense to ask for the probability of an event $A$ only if the conditions under which the event $A$ may or may not occur are specified and the value of the probability depends essentially on these conditions. In other words, every probability is in reality a conditional probability. This evident fact is somewhat obscured by the practice of omitting the explicit statement of the conditions if it is clear under which conditions the probability of an event is considered”
Hájek doesn’t take this argument to be sound, or else he wouldn’t have written his paper.

Rényi in fact seems not to either, because although he only ever considers conditional probabilities, he doesn’t conditionalize on information like “the pack is complete and well-shuffled”.

Instead, these premises are implicit in the model, rather than the antecedent of a conditional.
Hájek doesn’t take this argument to be sound, or else he wouldn’t have written his paper.

Rényi in fact seems not to either, because although he only ever considers conditional probabilities, he doesn’t conditionalize on information like “the pack is complete and well-shuffled”.

Instead, these premises are implicit in the model, rather than the antecedent of a conditional.

Thus, I take Rényi’s argument to show merely that every probability is relative to a model, not conditional.
My argument here is that for conditional probabilities, there is another relativization that must be considered.

In any real-world conditionalization, the antecedent arises as the answer to some earlier question, or the result of some earlier experiment.

The set of possible answers to this question, or results of the experiment, will give the partition to which conditional probability is relative.
Conditional probabilities must satisfy the integral equation, so if no function does, then not all conditional probabilities can exist.
Conditional probabilities must satisfy the integral equation, so if no function does, then not all conditional probabilities can exist.

If they exist, then whether or not conditional probabilities are basic, they must be relativized to a partition of the space.
Conditional probabilities must satisfy the integral equation, so if no function does, then not all conditional probabilities can exist.

If they exist, then whether or not conditional probabilities are basic, they must be relativized to a partition of the space.

In addition, they must (almost) agree with Kolmogorov’s elaboration, so they can be (almost) defined in terms of unconditional probabilities, assuming countable additivity.
Conditional probabilities must satisfy the integral equation, so if no function does, then not all conditional probabilities can exist.

If they exist, then whether or not conditional probabilities are basic, they must be relativized to a partition of the space.

In addition, they must (almost) agree with Kolmogorov’s elaboration, so they can be (almost) defined in terms of unconditional probabilities, assuming countable additivity.

Thus, there is little need to take conditional probability to be basic.
It is also reassuring to the naturalist that adopting this elaboration would put philosophers in harmony with the practice of mathematicians working in probability theory.

This analysis appears in Kolmogorov’s original book, and thus it is disingenuous to say that Kolmogorov was wrong when proposing the ratio analysis.

This three-place function is what conditional probability must (almost) be, if it exists.
One might try to repair a two-place function, noticing that the arguments above allowed for deviations from the required answers on sets of measure 0.

However, trying to come up with a two-place conditional probability function that satisfies all the requirements posed by rotating around different axes results in an ad hoc seeming function that is highly non-unique, and seems to require set-theoretic principles generally regarded to be false (described in my third appendix).
One might try to repair a two-place function, noticing that the arguments above allowed for deviations from the required answers on sets of measure 0.

However, trying to come up with a two-place conditional probability function that satisfies all the requirements posed by rotating around different axes results in an ad hoc seeming function that is highly non-unique, and seems to require set-theoretic principles generally regarded to be false (described in my third appendix).
Assuming countable additivity, the Radon-Nikodym theorem of real analysis says that there always exists a function $g_A$ satisfying the integral equation for $A$ and $E$.

However, there are many such functions, differing only on sets of measure 0.

Some method must still be found for choosing representatives of this equivalence class.

Kolmogorov’s elaboration consists of taking such a function for each pair $A$ and $E$ (where $E$ is a partition of the space including $B$ as one of its elements), and making them into one three-place function.
Conditions for a Regular Conditional Distribution

- $P(\cdot|B, E)$ should be a probability function
- $P(A|\cdot, E)$ should be $E$-measurable
- $P(A|B, \cdot)$ should be defined for all $E$ such that $B$ is $E$-measurable and $E$ is a sub-$\sigma$-field of the original space, not just when $E$ is generated by a partition of the space including $B$

- Seidenfeld et al call such a function a regular conditional distribution
Seidenfeld, Schervish, and Kadane

- Seidenfeld et al show that for certain RCDs, there will always be certain points $w$ and partitions $E$ such that $P(\{w\} | \{w\}, E) = 0$

- In addition, for certain spaces and sub-$\sigma$-fields, there are functions satisfying the integral equation such that almost every point $w$ has this property!

- This is clearly a problem
Some of the spaces in which this problem can occur almost everywhere simultaneously have other RCDs in which the problem is more contained.

They advocate the "finitely additive conditional probabilities" of Lester Dubins, since they are a two-place function that satisfies their desiderata.

But these functions don’t always satisfy the integral equation (and thus the reflection principle), so I don’t think they are an acceptable solution.
Thus, Seidenfeld et al use some spaces to show that RCDs can’t be conditional probabilities.

I show that if all the conditional probabilities exist, then they must be RCDs.

In some particular spaces, I show that these are necessarily three-place functions and not merely two-place ones.

Thus, unless one wants to sacrifice some of our conditions, not all conditional probabilities can exist in all spaces.