Title: Motivating outcome independence: locality versus sufficiency

Date: Jun 10, 2008 04:00 PM

URL: http://pirsa.org/08060190

Abstract: It is well known that the derivation of the Bell Inequality rests on two major assumptions, usually called outcome independence and parameter independence. Parameter independence seems to have a straightforward motivation: it expresses a non-signalling requirement between space-like separated sites and is thus motivated by locality. The status of outcome independence is much les clear. Many authors have argued that this assumption too expresses a locality requirement, in the form of a \'screening off\' condition. I will argue that the assumption also admits of an entirely different interpretation, suggested by the concept of sufficiency in the general theory of statistical inference. In this view, the assumption of outcome independence can be explained as expressing the idea that the specification of the hidden variable is sufficient, i.e. it exhausts all the relevant statistical information about the measurement outcomes. In this view, the assumption has no roots in locality at all. Rather, I would claim, it stems from the assumption that there exists such an exhaustive state description in our putative hidden variable theories.

Motivating outcome independence: locality versus sufficiency

Jos Uffink

June 10 2008

OUTLINE

1. What is at stake in the Bell Inequalities?

2. The usual assumptions in the derivation: (mainly) Parameter Independence and Outcome Independence.

3. The criterion of "Sufficiency" in statistical inference

4. Claim: this criterion provides an alternative motivation for Outcome Independence that does not rely on locality.

1. WHAT IS AT STAKE IN THE BELL INEQUALITIES?

Which kind of theories are ruled out through the violation of these inequalities by QM? In other words: what is the motivation behind the assumptions to derive the Bell Inequalities?

- "Received view": Conjunction of locality and realism

- "Alternative view": Just locality. (Stapp, Eberhard, Maudlin, Goldstein, Norsen, e.a.)

- Leggett: violations of BI "has relatively little to do with locality and much to do with objectivity"

The debate goes on...

1. WHAT IS AT STAKE IN THE BELL INEQUALITIES?

Which kind of theories are ruled out through the violation of these inequalities by QM? In other words: what is the motivation behind the assumptions to derive the Bell Inequalities?

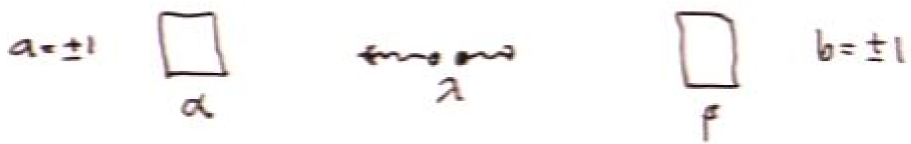
- "Received view": Conjunction of locality and realism
- "Alternative view": Just locality. (Stapp, Eberhard, Maudlin, Goldstein, Norsen, e.a.)
- Leggett: violations of BI "has relatively little to do with locality and much to do with objectivity"

The debate goes on...

My claim that an important source of inspiration has been overlooked in this debate. I will argue that the concept of sufficiency, first formulated by R.A. Fisher in 1922 in the context of mathematical statistics can throw a relevant light on this debate. As a demonstration of this relevance, I will argue that this concept of suffiency can be used to pinpoint and clarify the nature of (implicit or explicit) assumptions of "outcome independence" or "completeness" in the Bell Inequalities, and that this also underlines how much the argument relies on substantial realistic assumptions.

ASSUMPTIONS TO DERIVE BELL INEQUALITIES

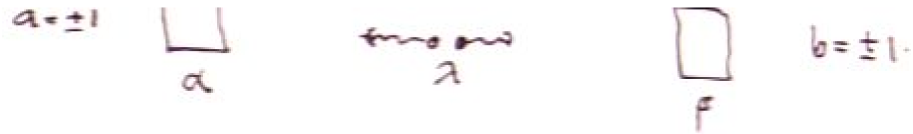
EPR experiment:



Assumptions:

1. There is a variable (or set of variables) $\lambda \in \Lambda$ characterizing the state of the particle pair. They are mutually exclusive and Λ is a measurable space.

2. The model specifies probabilities of obtaining the outcomes (a, b) for each choice of



Assumptions:

1. There is a variable (or set of variables) $\lambda \in \Lambda$ characterizing the state of the particle pair. They are mutually exclusive and Λ is a measurable space.

2. The model specifies probabilities of obtaining the outcomes (a, b) for each choice of the orientations α, β, if the pair is produced in the state λ:

$$p_{\alpha\beta}(a, b|\lambda) \qquad (1)$$

The empirically accessible probabilities of the outcomes are obtained by averaging over some probability density $\rho(\lambda)$:

$$p_{\alpha\beta}(a, b) = \int d\lambda p_{\alpha\beta}(a, b|\lambda)\rho_{\alpha,\beta}(\lambda) \qquad (2)$$

3. *Source Independence:*

the orientations α, β, if the pair is produced in the state λ:

$$p_{\alpha\beta}(a, b|\lambda) \tag{1}$$

The empirically accessible probabilities of the outcomes are obtained by averaging over some probability density $\rho(\lambda)$:

$$p_{\alpha\beta}(a, b) = \int d\lambda p_{\alpha\beta}(a, b|\lambda)\rho_{\alpha,\beta}(\lambda) \tag{2}$$

3. *Source Independence:*

$$\rho_{\alpha\beta}(\lambda) = \rho(\lambda) \tag{3}$$

4. *Outcome independence*

$$p_{\alpha\beta}(ab|\lambda) = p_{\alpha\beta}(a|\lambda)p_{\alpha\beta}(b|\lambda) \tag{4}$$

5. *Parameter Independence:*

$$\begin{aligned} p_{\alpha\beta}(a|\lambda) &= p_{\alpha}(a|\lambda) \\ p_{\alpha\beta}(b|\lambda) &= p_{\beta}(b|\lambda) \end{aligned} \tag{5}$$

the orientations α, β, if the pair is produced in the state λ:

$$p_{\alpha\beta}(a, b|\lambda) \qquad (1)$$

The empirically accessible probabilities of the outcomes are obtained by averaging over some probability density $\rho(\lambda)$:

$$p_{\alpha\beta}(a, b) = \int d\lambda p_{\alpha\beta}(a, b|\lambda)\rho_{\alpha,\beta}(\lambda) \qquad (2)$$

3. *Source Independence:*

$$\rho_{\alpha\beta}(\lambda) = \rho(\lambda) \qquad (3)$$

4. *Outcome independence*

$$p_{\alpha\beta}(ab|\lambda) = p_{\alpha\beta}(a|\lambda)p_{\alpha\beta}(b|\lambda) \qquad (4)$$

5. *Parameter Independence:*

$$p_{\alpha\beta}(a|\lambda) = p_{\alpha}(a|\lambda)$$
$$p_{\alpha\beta}(b|\lambda) = p_{\beta}(b|\lambda) \qquad (5)$$

Together, (4,5) imply the *factorization* condition:

$$p_{\alpha,\beta}(ab|\lambda) = p_\alpha(a|\lambda)p_\beta(b|\lambda) \qquad (6)$$

which imply the Bell Inequalities in a straightforward fashion. Bell called (6) "local causality".

REMARKS

1. Why distinguish between parameters and outcomes? Outcomes a and b, as well as the hidden variables λ are random variables, and figure as arguments of a joint probability distribution $p_{\alpha,\beta}(a,b,\lambda) = p_{\alpha,\beta}(a,b|\lambda)\rho(\lambda)$. The measurement settings α,β appear as parameters or labels of this probability function, not as arguments. Why not treat parameters and outcomes on the same footing, by writing

$$p_{\alpha,\beta}(a,b|\lambda) = p(a,b|\lambda\alpha\beta)? \qquad (7)$$

Together, (4,5) imply the *factorization* condition:

$$p_{\alpha,\beta}(ab|\lambda) = p_\alpha(a|\lambda)p_\beta(b|\lambda) \qquad (6)$$

which imply the Bell Inequalities in a straightforward fashion. Bell called (6) "local causality".

REMARKS

1. Why distinguish between parameters and outcomes? Outcomes a and b, as well as the hidden variables λ are random variables, and figure as arguments of a joint probability distribution $p_{\alpha,\beta}(a, b, \lambda) = p_{\alpha,\beta}(a, b|\lambda)\rho(\lambda)$. The measurement settings α, β appear as parameters or labels of this probability function, not as arguments. Why not treat parameters and outcomes on the same footing, by writing

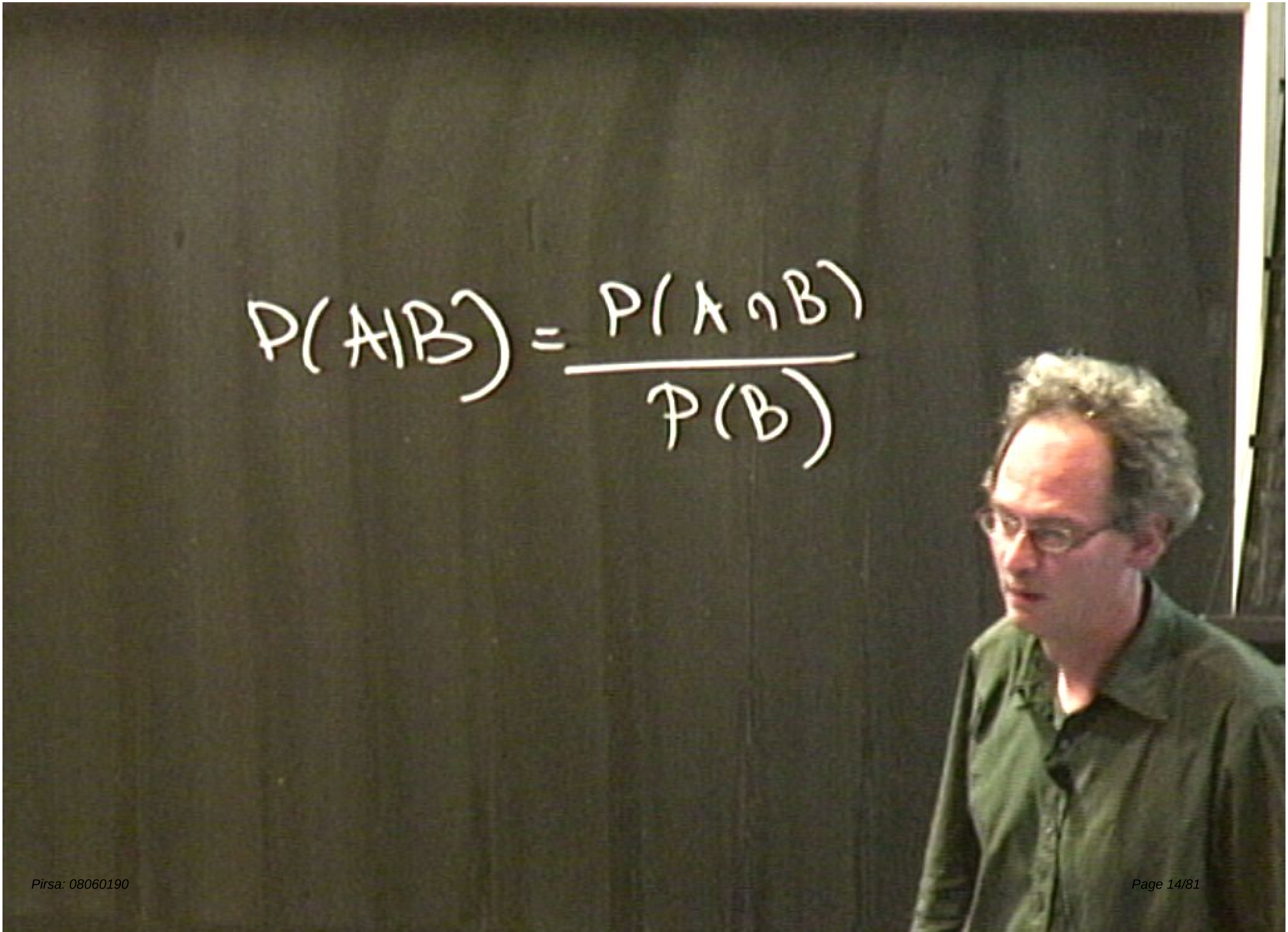
$$p_{\alpha,\beta}(a, b|\lambda) = p(a, b|\lambda\alpha\beta)? \qquad (7)$$

1. Why distinguish between parameters and outcomes? Outcomes a and b, as well as the hidden variables λ are random variables, and figure as arguments of a joint probability distribution $p_{\alpha,\beta}(a,b,\lambda) = p_{\alpha,\beta}(a,b|\lambda)\rho(\lambda)$. The measurement settings α,β appear as parameters or labels of this probability function, not as arguments. Why not treat parameters and outcomes on the same footing, by writing

$$p_{\alpha,\beta}(a,b|\lambda) = p(a,b|\lambda\alpha\beta)? \quad (7)$$

But treating α,β as conditioning arguments in a probability distribution, implies that a probability distribution over their possible values is defined. In other words, this means that the model would have to specify how probable it is that Alice will choose one setting α_1 rather than α_2, etc.

2. The first assumption is that there is a description of the particle pair in terms of some physical state $\lambda \in \Lambda$. this is motivated by realism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beable"


$P(A|B) = \frac{P(A \cap B)}{P(B)}$

But treating α, β as conditioning arguments in a probability distribution, implies that a probability distribution over their possible values is defined. In other words, this means that the model would have to specify how probable it is that Alice will choose one setting α_1 rather than α_2, etc.

2. The first assumption is that there is a description of the particle pair in terms of some physical state $\lambda \in \Lambda$. this is motivated by realism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beables".

3. Often a further requirement on λ is mentioned: that it provide a *complete* (or *full* or *total* or *exhaustive*) specification of the state in question. As we will see later on, some such requirement will be crucial to our discussion.

But treating α, β as conditioning arguments in a probability distribution, implies that a probability distribution over their possible values is defined. In other words, this means that the model would have to specify how probable it is that Alice will choose one setting α_1 rather than α_2, etc.

2. The first assumption is that there is a description of the particle pair in terms of some physical state $\lambda \in \Lambda$. this is motivated by realism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beables".

3. Often a further requirement on λ is mentioned: that it provide a *complete* (or *full* or *total* or *exhaustive*) specification of the state in question. As we will see later on, some such requirement will be crucial to our discussion.

Together, (4,5) imply the *factorization* condition:

$$p_{\alpha,\beta}(ab|\lambda) = p_{\alpha}(a|\lambda)p_{\beta}(b|\lambda) \qquad (6)$$

which imply the Bell Inequalities in a straightforward fashion. Bell called (6) "local causality".

REMARKS

1. Why distinguish between parameters and outcomes? Outcomes a and b, as well as the hidden variables λ are random variables, and figure as arguments of a joint probability distribution $p_{\alpha,\beta}(a, b, \lambda) = p_{\alpha,\beta}(a, b|\lambda)\rho(\lambda)$. The measurement settings α, β appear as parameters or labels of this probability function, not as arguments. Why not treat parameters and outcomes on the same footing, by writing

$$p_{\alpha,\beta}(a, b|\lambda) = p(a, b|\lambda\alpha\beta)? \qquad (7)$$

Together, (4,5) imply the *factorization* condition:

$$p_{\alpha,\beta}(ab|\lambda) = p_\alpha(a|\lambda)p_\beta(b|\lambda) \qquad (6)$$

which imply the Bell Inequalities in a straightforward fashion. Bell called (6) "local causality".

REMARKS

1. Why distinguish between parameters and outcomes? Outcomes a and b, as well as the hidden variables λ are random variables, and figure as arguments of a joint probability distribution $p_{\alpha,\beta}(a,b,\lambda) = p_{\alpha,\beta}(a,b|\lambda)\rho(\lambda)$. The measurement settings α, β appear as parameters or labels of this probability function, not as arguments. Why not treat parameters and outcomes on the same footing, by writing

$$p_{\alpha,\beta}(a,b|\lambda) = p(a,b|\lambda\alpha\beta)? \qquad (7)$$

But treating α, β as conditioning arguments in a probability distribution, implies that a probability distribution over their possible values is defined. In other words, this means that the model would have to specify how probable it is that Alice will choose one setting α_1 rather than α_2, etc.

2. The first assumption is that there is a description of the particle pair in terms of some physical state $\lambda \in \Lambda$. this is motivated by realism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beables".

3. Often a further requirement on λ is mentioned: that it provide a *complete* (or *full* or *total* or *exhaustive*) specification of the state in question. As we will see later on, some such requirement will be crucial to our discussion.

is that Alice will choose one setting α_1 rather than α_2, etc.

2. The first assumption is that there is a description of the particle pair in terms of some physical state $\lambda \in \Lambda$. this is motivated by realism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beables".

3. Often a further requirement on λ is mentioned: that it provide a *complete* (or *full* or *total* or *exhaustive*) specification of the state in question. As we will see later on, some such requirement will be crucial to our discussion.

ism: λ is independent of what observations we choose to make on the system. (the 'measurement context'). To make this point explicit, Bell coined the expression "beables".

3. Often a further requirement on λ is mentioned: that it provide a *complete* (or *full* or *total* or *exhaustive*) specification of the state in question. As we will see later on, some such requirement will be crucial to our discussion.

However, one can only rarely find that this idea is spelt out in any detail, let alone translated into a formal and exact condition.

At first sight, it might seem that the requirement is simple enough, and not in need of further explication. The trouble arises when we realize that our argument is supposed to cover a general class of models or theories. And each of these theories might postulate a different kind of reality, i.e. a different collection of properties and or relations of the system. Now for any putative state space Λ, we can always envisage a class of possible worlds such that a state $\lambda \in \Lambda$ does provide a *complete* description of the system in the corresponding world;

a general class of models or theories. And each of these theories might postulate a different kind of reality, i.e. a different collection of properties and or relations of the system. Now for any putative state space Λ, we can always envisage a class of possible worlds such that a state $\lambda \in \Lambda$ does provide a *complete* description of the system in the corresponding world; simply because that world is defined as one in which there are only such properties and relations for the system as contained in the state λ. At the same time, any such space Λ might also be considered as *incomplete*, since it is always

possible to regard that space as the quotient obtained from a a partition over some more refined space Λ', such that each $\lambda \in \Lambda$ would correspond to some element in the partition over Λ', i.e. as a coarse grained state in the latter.

In other words, if we take 'complete' to mean that a state description λ provide all the properties that the system has, λ can be seen as complete, or incomplete, as long as we allow arbitrary theories or models, and so it is not straightforward to give this intuitively idea a clear role in the mathematical formulation of possible to regard that space as the quotient obtained from a a partition over some more refined space Λ', such that each $\lambda \in \Lambda$ would correspond to some element in the partition over Λ', i.e. as a coarse grained state in the latter.

In other words, if we take 'complete' to mean that a state description λ provide all the properties that the system has, λ can be seen as complete, or incomplete, as long as we allow arbitrary theories or models, and so it is not straightforward to give this intuitively idea a clear role in the mathematical formulation of the argument. For the moment, I will therefore shelve this problem and ignore the question of whether the state description is complete.

4. The motivation for Source Independence is relatively straightforward. Often, one invokes the free will of the experimenter to choose

complete, or incomplete, as long as we allow arbitrary theories or models, and so it is not straightforward to give this intuitively idea a clear role in the mathematical formulation of the argument. For the moment, I will therefore shelve this problem and ignore the question of whether the state description is complete.

4. The motivation for Source Independence is relatively straightforward. Often, one invokes the free will of the experimenter to choose

these settings in an arbitrary manner, even after the particles have left the source.

5. The motivation for Parameter Independence is also relatively straightforward: it expresses a demand that the model does not allow superluminal signalling, even in principle.

MOTIVATIONS FOR OUTCOME INDEPENDENCE

Condition (4) was introduced by Jarrett as a condition of *completeness*.

these settings in an arbitrary manner, even after the particles have left the source.

5. The motivation for Parameter Independence is also relatively straightforward: it expresses a demand that the model does not allow superluminal signalling, even in principle.

MOTIVATIONS FOR OUTCOME INDEPENDENCE

Condition (4) was introduced by Jarrett as a condition of *completeness*.

> " If all information about particle 1 is included in the state description [λ] then knowledge of the outcome b provides only redundant information about particle 1."

these settings in an arbitrary manner, even after the particles have left the source.

5. The motivation for Parameter Independence is also relatively straightforward: it expresses a demand that the model does not allow superluminal signalling, even in principle.

MOTIVATIONS FOR OUTCOME INDEPENDENCE

Condition (4) was introduced by Jarrett as a condition of *completeness.*

> " If all information about particle 1 is included in the state description [λ] then knowledge of the outcome b provides only redundant information about particle 1."

5. The motivation for Parameter Independence is also relatively straightforward: it expresses a demand that the model does not allow superluminal signalling, even in principle.

MOTIVATIONS FOR OUTCOME INDEPENDENCE

Condition (4) was introduced by Jarrett as a condition of *completeness*.

> " If all information about particle 1 is included in the state description [λ] then knowledge of the outcome b provides only redundant information about particle 1."

presses a demand that the model does not allow superluminal signalling, even in principle.

MOTIVATIONS FOR OUTCOME INDEPENDENCE

Condition (4) was introduced by Jarrett as a condition of *completeness*.

> " If all information about particle 1 is included in the state description [λ] then knowledge of the outcome b provides only redundant information about particle 1."

Shimony, in an effort to be more neutral called it "parameter independence" and argued that the condition could be explained be an appeal to locality, albeit different from superluminal signalling.

Others have argued that Outcome Independence can motivated by (some variant of) Reichenbach's Principle of the Common Cause or Hellmann's Stochastic Einstein Locality. (cf. Butterfield)

Shimony, in an effort to be more neutral called it "parameter independence" and argued that the condition could be explained be an appeal to locality, albeit different from superluminal signalling.

Others have argued that Outcome Independence can motivated by (some variant of) Reichenbach's Principle of the Common Cause or Hellmann's Stochastic Einstein Locality. (cf. Butterfield)

Elby, Brown and Foster claim that "Jarett completeness follows from natural assumptions about locality and causality [?, p. 977]". Berkovitz : "Violations of Outcome independence involve some type of holism, non-seperability and/or passion at a distance that may be posible to reconcile with relativity. [?]"

Going even further, Norsen argues " The whole motivation of Jarrett's project . . . is based on fundamental confusion."

outcome

Shimony, in an effort to be more neutral called it "parameter independence" and argued that the condition could be explained be an appeal to locality, albeit different from superluminal signalling.

Others have argued that Outcome Independence can motivated by (some variant of) Reichenbach's Principle of the Common Cause or Hellmann's Stochastic Einstein Locality. (cf. Butterfield)

Elby, Brown and Foster claim that " Jarett completeness follows from natural assumptions about locality and causality [?, p. 977]". Berkovitz : "Violations of Outcome independence involve some type of holism, non-seperability and/or passion at a distance that may be posible to reconcile with relativity. [?]"

Going even further, Norsen argues " The whole motivation of Jarrett's project . . . is based on fundamental confusion."

Shimony, in an effort to be more neutral called it "parameter independence" and argued that the condition could be explained be an appeal to locality, albeit different from superluminal signalling.

Others have argued that Outcome Independence can motivated by (some variant of) Reichenbach's Principle of the Common Cause or Hellmann's Stochastic Einstein Locality. (cf. Butterfield)

Elby, Brown and Foster claim that " Jarett completeness follows from natural assumptions about locality and causality [?, p. 977]". Berkovitz : "Violations of Outcome independence involve some type of holism, non-seperability and/or passion at a distance that may be posible to reconcile with relativity. [?]"

Going even further, Norsen argues " The whole motivation of Jarrett's project . . . is based on fundamental confusion."

I will argue that one may look at the condition of outcome independence as providing the flesh and bone of the realist supposition by specifying that λ (together with α and β) should be complete, or as I prefer to say *sufficient* for the outcomes a and b.

Caveat:

Consider a theory that posits, that, apart from the hidden variable λ, there is also a deeper level variable μ and that all the probabilities $p_{\alpha,\beta}(a,b|\lambda)$ are actually averages over the additional variable. Thus:

$$p_{\alpha,\beta}(a,b|\lambda) = \int p_{\alpha,\beta}(a,b|\lambda,\mu)\rho(\mu|\lambda) \qquad (8)$$

If Outcome Independence (or Parameter Independence) hold at one level, must they also hold at the other level?

No! Examples:

dition of outcome independence as providing the flesh and bone of the realist supposition by specifying that λ (together with α and β) should be complete, or as I prefer to say *sufficient* for the outcomes a and b.

Caveat:

Consider a theory that posits, that, apart from the hidden variable λ, there is also a deeper level variable μ and that all the probabilities $p_{\alpha,\beta}(a, b|\lambda)$ are actually averages over the additional variable. Thus:

$$p_{\alpha,\beta}(a, b|\lambda) = \int p_{\alpha,\beta}(a, b|\lambda, \mu)\rho(\mu|\lambda) \qquad (8)$$

If Outcome Independence (or Parameter Independence) hold at one level, must they also hold at the other level?

No! Examples:

- Orthodox QM: $\lambda = |\Psi\rangle_{\text{singlet}}$.

  PI holds; OI fails.

- De Broglie-Bohm. $\lambda = (|\Psi\rangle_{\text{singlet}}, \vec{x_1}, \vec{x_2})$.

  PI fails, OI holds.

- Leggett's (deteministic) model $\lambda = (\mu, \vec{u}, \vec{v})$
  PI fails, PI holds.

  Average over μ. PI holds, OI fails.

- Orthodox QM: $\lambda = |\Psi\rangle_{\text{singlet}}$.

  PI holds; OI fails.

- De Broglie-Bohm. $\lambda = (|\Psi\rangle_{\text{singlet}}, \vec{x_1}, \vec{x_2})$.

  PI fails, OI holds.

- Leggett's (deteministic) model $\lambda = (\mu, \vec{u}, \vec{v})$
  PI fails, OI holds.

  Average over μ. PI holds, OI fails.

SUFFICIENCY IN STATISTICAL INFERENCE

Sufficiency was developed in the theory of statistical estimation by R.A. Fisher (1922).

Suppose we have some probabilistic experiment with possible outcomes $x \in X$ and a family of probability distributions p_θ, $\theta \in \Theta$, providing candidate description for the experiment.

Here, Θ represents some arbitrary index set. The problem is to infer which distribution - i.e. which θ would provide a best "fit" for the experiment on the basis of recorded outcomes.

Consider identically distributed repeated trials with outcomes $(x_1, \ldots, x_n)$ and

$$p_\theta(x_1, \ldots, x_n) = \prod_{i=1}^{n} p_\theta(x_i) \qquad (9)$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \ldots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \mapsto \Theta$ such that $\tau(x_1, \ldots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \cdots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \ldots x_n) &= c_1 \\ &\vdots \\ y_n(x_1, \ldots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \cdots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \ldots x_n)$ can be equivalently represented as $y_1, \cdots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\tilde{p}_\theta(y_1, \ldots y_n) = p(x_1, \ldots x_n) \left| \frac{\partial x_i}{\partial y_j} \right| \tag{11}$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \dots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \longmapsto \Theta$ such that $\tau(x_1, \dots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \cdots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \dots x_n) &= c_1 \\ &\vdots \\ y_n(x_1, \dots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \cdots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \dots x_n)$ can be equivalently represented as $y_1, \cdots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\tilde{p}_\theta(y_1, \cdots y_n) = p(x_1, \dots x_n) \left| \frac{\partial x_i}{\partial y_j} \right| \tag{11}$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \dots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \mapsto \Theta$ such that $\tau(x_1, \dots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \cdots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \dots x_n) &= c_1 \\ &\vdots \\ y_n(x_1, \dots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \cdots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \dots x_n)$ can be equivalently represented as $y_1, \cdots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\hat{p}_\theta(y_1, \dots y_n) = p(x_1, \dots x_n) |\frac{\partial x_i}{\partial y_j}| \tag{11}$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \dots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \mapsto \Theta$ such that $\tau(x_1, \dots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \cdots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \dots x_n) &= c_1 \\ \vdots &\quad \vdots \\ y_n(x_1, \dots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \cdots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \dots x_n)$ can be equivalently represented as $y_1, \dots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\hat{p}_\theta(y_1, \dots y_n) = p(x_1, \dots x_n) \left| \frac{\partial x_i}{\partial y_j} \right| \tag{11}$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \ldots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \mapsto \Theta$ such that $\tau(x_1, \ldots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \cdots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \ldots x_n) &= c_1 \\ &\vdots \\ y_n(x_1, \ldots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \cdots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \ldots x_n)$ can be equivalently represented as $y_1, \cdots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\hat{p}_\theta(y_1, \cdots y_n) = p(x_1, \cdots x_n) \left| \frac{\partial x_i}{\partial y_j} \right| \tag{11}$$

The goal now becomes to make an inference about θ on the basis of the outcomes $(x_1, \ldots x_n)$.

In estimation theory, one designs a function (estimator) $\tau : X^n \mapsto \Theta$ such that $\tau(x_1, \ldots, x_n)$ would represent the best estimate of θ.

Consider n independent functions $(y_1, \ldots y_n)$ defined on X^n, such that the equations

$$\begin{aligned} y_1(x_1, \ldots x_n) &= c_1 \\ \vdots \quad & \quad \vdots \\ y_n(x_1, \ldots x_n) &= c_n \end{aligned} \tag{10}$$

always have a unique solution. (Think of $(y_1, \ldots y_n)$ as an alternative coordinate system for the points in X^n.) The recorded outcomes $(x_1 \ldots x_n)$ can be equivalently represented as $y_1, \ldots y_n)$. The probability distribution can be transformed to the alternative coordinates:

$$\tilde{p}_\theta(y_1, \ldots y_n) = p(x_1, \ldots x_n) |\frac{\partial x_i}{\partial y_j}| \tag{11}$$

where the last factor represents the Jacobian of the transformation.

Now suppose that $\tilde{p}_\theta$ has the following form

$$\tilde{p}_\theta(y_1, \cdots y_n) = f_\theta(y_1) g(y_1, \cdots y_n) \qquad (12)$$

In that case, the function y_1 is said to be *sufficient* for θ.

Note that if y_1 out of the set $(y_1, \cdots y_n)$ is sufficient for θ, then, the same will hold for an alternative set $y_1, \tilde{y_2}, \cdots \tilde{y}_n$, provided that this is also a regular coordinate system.

In Fisher's own words, the criterion of suficiency is:

> That the statistic chosen should summarize the whole of the relevant information supplied by the sample. [. . .]

... then, the same will hold for an alternative set $y_1, \bar{y}_2, \ldots, \bar{y}_n$, provided that this is also a regular coordinate system.

In Fisher's own words, the criterion of sufficiency is:

> That the statistic chosen should summarize the whole of the relevant information supplied by the sample. [...]

> In mathematical language we may interpret this statement by saying that if θ is the parameter to be estimated, y_1 a statistic which contains the whole of the information as to the value of θ which the sample supplies, and y_2 any other statistic, then the surface of distribution of pairs of values y_1 and y_2 for a given value of θ is such that for a given value of y_1, the distribution of y_2 does not involve θ. In other words, when y_1 is known, knowledge of the value of y_2 throws no further light upon the value of θ. [?, p. 317]

Example: the normal distribution:

provides a sufficient statistic for θ. In general, however, the class of probability distributions for which a sufficient statistic exists is severely limited; the Pitman-Koopman theorem implies they exist only for the exponential family.

Note that notions of locality or causation can be kept safely on the bench in this approach.

Note: Fisher talk about "information", that is "contained" in a sample, and is "about" something. To say that y_1 is sufficient is qualified by saying that it is sufficient for a purpose (inferring the value of θ; it may wel be insufficient for other purposes!) and relative a to a class, namely all other statistics of the same sample, i.e. all other functions of the outcome space X^n.

Bayesian inference

The above approach used the orthodox statistical inference, in which parameters are kept strictly distinct from outcomes. At first sight, this dichotomy corresponds nicely to the same dichotomy adopted in the discussion above. However, in the context of the Bell's theorem we are not aiming at making an inference about α or β, but rather about outcomes a, b. Therefore we will have to change the perspective somewhat.

There is an alternative approach to statistical inference, the Bayesian approach, that does not rely on such a strict division between events and parameters.

Bayesian statistical inference assumes the existence of a so-called prior probability distribution over the parameter θ. Furthermore, the probability distributions $p_\theta(x)$ are now reinter-

The above approach used the orthodox statistical inference, in which parameters are kept strictly distinct from outcomes. At first sight, this dichotomy corresponds nicely to the same dichotomy adopted in the discussion above. However, in the context of the Bell's theorem we are not aiming at making an inference about α or β, but rather about outcomes a, b. Therefore we will have to change the perspective somewhat.

There is an alternative approach to statistical inference, the Bayesian approach, that does not rely on such a strict division between events and parameters.

Bayesian statistical inference assumes the existence of a so-called prior probability distribution over the parameter θ. Furthermore, the probability distributions $p_\theta(x)$ are now reinterpreted as conditional distributions

$$p_\theta(x) = p(x|\theta) \tag{15}$$

The above approach used the orthodox statistical inference, in which parameters are kept strictly distinct from outcomes. At first sight, this dichotomy corresponds nicely to the same dichotomy adopted in the discussion above. However, in the context of the Bell's theorem we are not aiming at making an inference about α or β, but rather about outcomes a, b. Therefore we will have to change the perspective somewhat.

There is an alternative approach to statistical inference, the Bayesian approach, that does not rely on such a strict division between events and parameters.

Bayesian statistical inference assumes the existence of a so-called prior probability distribution over the parameter θ. Furthermore, the probability distributions $p_\theta(x)$ are now reinterpreted as conditional distributions

$$p_\theta(x) = p(x|\theta) \tag{15}$$

The above approach used the orthodox statistical inference, in which parameters are kept strictly distinct from outcomes. At first sight, this dichotomy corresponds nicely to the same dichotomy adopted in the discussion above. However, in the context of the Bell's theorem we are not aiming at making an inference about α or β, but rather about outcomes a, b. Therefore we will have to change the perspective somewhat.

There is an alternative approach to statistical inference, the Bayesian approach, that does not rely on such a strict division between events and parameters.

Bayesian statistical inference assumes the existence of a so-called prior probability distribution over the parameter θ. Furthermore, the probability distributions $p_\theta(x)$ are now reinterpreted as conditional distributions

$$p_\theta(x) = p(x|\theta) \qquad (15)$$

Given these two assumptions, it is possible to provide a so-called posterior probability distribution by means of Bayes' theorem, i.e.

$$p(\theta|x) = \frac{p(x|\theta)\rho(\theta)}{\int d\theta p(x|\theta)\rho(\theta)} \qquad (16)$$

Extending this to the case of multiple, independent and identically distributed trials one obtains:

$$p(\theta|x_1, \ldots x_n) = \frac{\prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)}{\int \prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)d\theta} \qquad (17)$$

In the Bayesian approach, the goal of statistical inference is to report this posterior probability distribution.

Note that this viewpoint does not necessarily presuppose a subjective interpretation of probability.

Given these two assumptions, it is possible to provide a so-called posterior probability distribution by means of Bayes' theorem, i.e.

$$p(\theta|x) = \frac{p(x|\theta)\rho(\theta)}{\int d\theta p(x|\theta)\rho(\theta)} \qquad (16)$$

Extending t[illegible] to the case of multiple, independent an[illegible]dentically distributed trials one obtains:

$$p(\theta|x_1, \ldots, x_n) = \frac{\prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)}{\int \prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)d\theta} \qquad (17)$$

In the Ba[illegible]an approach, the goal of statistical infere[illegible] is to report this posterior probability dis[illegible]tion.

Note th[illegible]iewpoint does not necessarily presu[illegible]tive interpretation of probability[illegible]

Given these two assumptions, it is possible to provide a so-called posterior probability distribution by means of Bayes' theorem, i.e.

$$p(\theta|x) = \frac{p(x|\theta)\rho(\theta)}{\int d\theta p(x|\theta)\rho(\theta)} \qquad (16)$$

Extending this to the case of multiple, independent and identically distributed trials one obtains:

$$p(\theta|x_1, \ldots x_n) = \frac{\prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)}{\int \prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)d\theta} \qquad (17)$$

In the Bayesian approach, the goal of statistical inference is to report this posterior probability distribution.

Note that this viewpoint does not necessarily presuppose a subjective interpretation of probability.

Given these two assumptions, it is possible to provide a so-called posterior probability distribution by means of Bayes' theorem, i.e.

$$p(\theta|x) = \frac{p(x|\theta)\rho(\theta)}{\int d\theta p(x|\theta)\rho(\theta)} \quad (16)$$

Extending this to the case of multiple, independent and identically distributed trials one obtains:

$$p(\theta|x_1, \ldots x_n) = \frac{\prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)}{\int \prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)d\theta} \quad (17)$$

In the Bayesian approach, the goal of statistical inference is to report this posterior probability distribution.

Note that this viewpoint does not necessarily presuppose a subjective interpretation of probability.

Given these two assumptions, it is possible to provide a so-called posterior probability distribution by means of Bayes' theorem, i.e.

$$p(\theta|x) = \frac{p(x|\theta)\rho(\theta)}{\int d\theta p(x|\theta)\rho(\theta)} \qquad (16)$$

Extending this to the case of multiple, independent and identically distributed trials one obtains:

$$p(\theta|x_1, \ldots x_n) = \frac{\prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)}{\int \prod_{i=1}^{n} p_\theta(x_i)\rho(\theta)d\theta} \qquad (17)$$

In the Bayesian approach, the goal of statistical inference is to report this posterior probability distribution.

Note that this viewpoint does not necessarily presuppose a subjective interpretation of probability.

Now, the sufficiency condition is equivalent to

$$p(\theta|y_1,\dots y_n) = p(\theta|y_1) \qquad (18)$$

Perhaps, this makes the under[illegible]g motivation of sufficiency even clearer: if [illegible]e above condition holds, then the proba[illegible]ty of θ, once y_1 is given, is not changed w[illegible]n the values of $y_2,\dots y_n$ are added. These a[illegible]ional functions of the data are irrelevant o[illegible]dundant for the purpose of asigning this po[illegible]rior probability.

It should also now be clea[illegible]ow this condition would fit in with the con[illegible]n of Outcome Independence. first of al[illegible]e will suppose the condition (18) to hold [illegible]ch of the probability distribution[illegible]e settings α, β. Secondly, λ w[illegible]. Now, obviously, λ is no[illegible]ion of the data, it is stil[illegible]om variable. It has a [illegible]will in general be hidden[illegible]bout

Now, the sufficiency condition is equivalent to

$$p(\theta|y_1, \ldots y_n) = p(\theta|y_1) \tag{18}$$

Perhaps, this makes the underlying motivation of sufficiency even clearer: if the above condition holds, then the probability of θ, once y_1 is given, is not changed when the values of $y_2, \ldots y_n$ are added. These additional functions of the data are irrelevant or redundant for the purpose of asigning this posterior probability.

It should also now be clear how this condition would fit in with the condition of Outcome Independence. first of all, we will suppose the condition (18) to hold for each of the probability distributions labeled by the settings α, β. Secondly, λ will play the role of y_1. Now, obviously, λ is not a datum nor a function of the data, it is still assumed to be a random variable. It has a value, and although it will in general be hidden for us we can reason about

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \qquad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \qquad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \qquad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \qquad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \qquad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \qquad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \quad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \quad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \qquad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \qquad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

the hypothetical case that we would know its value.

By now it will come as no surprise what the essential proposal is.

Make the assumption that for each α, β, the hidden variables λ, are sufficient for the prediction of a.

Outcome Independence

$$p_{\alpha\beta}(a|\lambda, b) = p_{\alpha\beta}(a|\lambda) \qquad (19)$$

says that b is redundant for this purpose.

Similarly, one can state

$$p_{\alpha\beta}(ab|\lambda, \mu) = p_{\alpha\beta}(ab|\lambda) \qquad (20)$$

for any 'deeper' specification of the hidden variables. This formalizes in what sense λ would be a 'complete' state.

4. DISCUSSION

First of all, I would claim that this paper shows that a reconsideration of the theory of statistical inference pays off in a discussion of the assumptions needed to derive the Bell Inequalities. The assumption of outcomes independence can be understood as an application of the criterion of sufficiency: one may understand the condition of outcome independence as stating that the hidden variable λ is sufficient for the prediction of the outcomes in each wing of the EPR experiment in the sense that the second outcome would only provide redundant information once λ is given. Of course this is very similar to the point of view taken by Jarrett, and, in this sense this paper can be seen as a vindication of Jarrett's view by connecting it to a venerable but entirely overlooked tradition in statistics.

However, this does not mean that the above motivation would be the only one, or indeed the most persuasive motivation for Outcome Independence. So let us consider the alternative views, i.e. that it is motivated by locality of causality.

For example, let us look at the principle of the common cause. This principle states that whenever there exists a probabilistic correlation between events a and b, i.e. $p(a,b) \neq p(a)p(b)$ there always exists an event C (in the past of A and B) such that

$$p(a,b|C) = p(a|C)p(b|C) \quad (21)$$

$$p(a,b|\neg C) = p(a\neg C)p(b|\neg C) \quad (22)$$

If such an event C exists, it is said to be the *common cause* of the correlation. Reichenbach's formulation suffers from various defects. Most importantly, for the present purposes, it

However, this does not mean that the above motivation would be the only one, or indeed the most persuasive motivation for Outcome Independence. So let us consider the alternative views, i.e. that it is motivated by locality of causality.

For example, let us look at the principle of the common cause. This principle states that whenever there exists a probabilistic correlation between events a and b, i.e. $p(a,b) \neq p(a)p(b)$ there always exists an event C (in the past of A and B) such that

$$p(a,b|C) = p(a|C)p(b|C) \qquad (21)$$

$$p(a,b|\neg C) = p(a\neg C)p(b|\neg C) \qquad (22)$$

If such an event C exists, it is said to be the *common cause* of the correlation. Reichenbach's formulation suffers from various defects. Most importantly, for the present purposes, it

For example, let us look at the principle of the common cause. This principle states that whenever there exists a probabilistic correlation between events a and b, i.e. $p(a,b) \neq p(a)p(b)$ there always exists an event C (in the past of A and B) such that

$$p(a,b|C) = p(a|C)p(b|C) \tag{21}$$

$$p(a,b|\neg C) = p(a|\neg C)p(b|\neg C) \tag{22}$$

If such an event C exists, it is said to be the common cause of the correlation. Reichenbach's formulation suffers from various defects. Most importantly, for the present purposes, it

For example, let us look at the principle of the common cause. This principle states that whenever there exists a probabilistic correlation between events a and b, i.e. $p(a,b) \neq p(a)p(b)$ there always exists an event C (in the past of A and B) such that

$$p(a,b|C) = p(a|C)p(b|C) \qquad (21)$$

$$p(a,b|\neg C) = p(a\neg C)p(b|\neg C) \qquad (22)$$

If such an event C exists, it is said to be the *common cause* of the correlation. Reichenbach's formulation suffers from various defects. Most importantly, for the present purposes, it

For example, let us look at the principle of the common cause. This principle states that whenever there exists a probabilistic correlation between events a and b, i.e. $p(a,b) \neq p(a)p(b)$ there always exists an event C (in the past of A and B) such that

$$p(a,b|C) = p(a|C)p(b|C) \quad (21)$$
$$p(a,b|\neg C) = p(a\neg C)p(b|\neg C) \quad (22)$$

If such an event C exists, it is said to be the *common cause* of the correlation. Reichenbach's formulation suffers from various defects. Most importantly, for the present purposes, it does not say anything about what would happen when we partition C into two (or more subevents).

For example, Anne and Bill live on opposite sides of Rainy Mountain. Anne finds that when her basement floods, Bill's basement is also disproportionally likely to flood. Or more precisely: $p(b|a) > p(b)$. However, one would not think of the flooding of Anne's basement causes the flooding of Bill's basement. Now suppose that r denotes the event that there does not say anything about what would happen when we partition C into two (or more subevents).

For example, Anne and Bill live on opposite sides of Rainy Mountain. Anne finds that when her basement floods, Bill's basement is also disproportionally likely to flood. Or more precisely: $p(b|a) > p(b)$. However, one would not think of the flooding of Anne's basement causes the flooding of Bill's basement. Now suppose that r denotes the event that there was a rainstorm on the previous night at rainy mountain. We find that

$$p(a, b|r) = p(a|r)p(b|r)$$

Thus, r seems to provide the common cause of the flooding of the basements.

But now suppose we divide the rainstorms into types: heavy rainstorms (r_h) and light rainstorms (r_l). And suppose that conditioning on the precise type of rainstorm, we find that the disappearance of the correlation is actually due to a mixture of positive correlations for heavy rainstorms and an anticorrelation fot light storms. Thus

$$p(ab|r_h) > p(a|r_h)p(b|r_h) \quad p(ab|r_l) < p(ar_l)p(b|r_l) \tag{23}$$

Would we still consider the event "rainstorm" as the common cause of the correlation between a and b? I don't think so. In any case, we would now be left with further correlations to explain. And when we explained them (e.g. by an even more precise description of the type of precipitation), we should worry that further description might reveal further correlations.

Again, this possibility is mathematically trivial: factorisation conditions such as

$$p(a,b|r) = p(a|r)p(b|r) \tag{24}$$

are not conserved under convex combinations. This means that the PCC as it stands is not suitable. This has not escaped other commentators. Thus, Elby, Brown and Forster add a significant condtion: ' it is when we build into r a *statistically complete* description of the rainfall, wind patterns, drainage condtions etc., then we expect that r is the Reichenbachian common cause of a and b. . . . In words: given a *full* description of the rainstorm, the probability of Bill's basement flooding will not be lowered or raised by the additional information that Anne's basement is flooded. Information about Anne's basement is redundant or irrelevant once we posess the precise weather information contained in r. [?, p. 981].

Thus, according to these authors, it is only when we amend Reichenbach's statement by the assumption that the conditioning event r is "statistically complete" that we can expect Reichenbach's principle to hold.

The question is now what these authors understand by 'statistically complete". Unfortunately, their definition involves a relation between theoretical predictions and observed frequencies. (This is unfortunate because one would like to keep the isue of whether a theory is complete and whether the theory is empirically adequate separated.) However, there is a corrolary of their definition that can be formulated without reference to observation. Consider a theory and an ensemble of systems prepared in exactly the same state, as characterized by this theory. If the theory (or the state description) is statistically complete, then there is no *conceivable* way to split the ensemble into subensembles with different measurement statistics.

But this is *exactly* the point to be argued. For if the factorisation (24) does not hold, then by sorting out members of an ensemble

The question is now what these authors understand by 'statistically complete". Unfortunately, their definition involves a relation between theoretical predictions and observed frequencies. (This is unfortunate because one would like to keep the isue of whether a theory is complete and whether the theory is empirically adequate separated.) However, there is a corrolary of their definition that can be formulated without reference to observation. Consider a theory and an ensemble of systems prepared in exactly the same state, as characterized by this theory. If the theory (or the state description) is statistically complete, then there is no *conceivable* way to split the ensemble into subensembles with different measurement statistics.

But this is *exactly* the point to be argued. For if the factorisation (24) does not hold, then by sorting out members of an ensemble

does not say anything about what would happen when we partition C into two (or more subevents).

For example, Anne and Bill live on opposite sides of Rainy Mountain. Anne finds that when her basement floods, Bill's basement is also disproportionally likely to flood. Or more precisely: $p(b|a) > p(b)$. However, one would not think of the flooding of Anne's basement causes the flooding of Bill's basement. Now suppose that r denotes the event that there was a rainstorm on the previous night at rainy mountain. We find that

$$p(a, b|r) = p(a|r)p(b|r)$$

Thus, r seems to provide the common cause of the flooding of the basements.

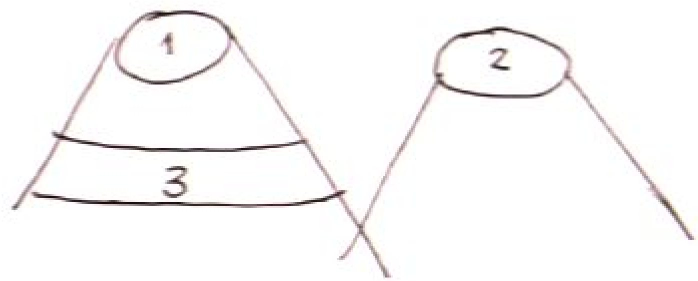
But now suppose we divide the rainstorms into types: heavy rainstorms (r_h) and light rainstorms (r_l). And suppose that conditioning

The question is now what these authors understand by 'statistically complete". Unfortunately, their definition involves a relation between theoretical predictions and observed frequencies. (This is unfortunate because one would like to keep the isue of whether a theory is complete and whether the theory is empirically adequate separated.) However, there is a corrolary of their definition that can be formulated without reference to observation. Consider a theory and an ensemble of systems prepared in exactly the same state, as characterized by this theory. If the theory (or the state description) is statistically complete, then there is no *conceivable* way to split the ensemble into subensembles with different measurement statistics.

But this is *exactly* the point to be argued. For if the factorisation (24) does not hold, then by sorting out members of an ensemble

Bell himself has never adopted the breaking-up of the factorization condition into two separate subconditions (parameter independence and outcome independence). For him, the factorization condition was a package deal ("local causality"). It has been argued (e.g. by Norsen) that this means that he did not think that a substantial assumption of realism or completeness was at stake. However, such an argument overlooks that Bell formulated this assumption of "local causality" for a particular kind of physical theories, namely a theory of beables.

Further, Bell stressed the importance of "completeness":



A theory is said to be locally causal if the probabilities atached to values of local beables in a space-time region 1 are unaltered by a specification of values of local beables in a space-like separated region 2 when what happens in the backward light cone is already *sufficiently* specified, for example by a full specification of local beables in a space-time region 3. It is important that region 3 completely shields off from 1 the overlap of the backward light cones

If the probabilities attached to values of local beables in a space-time region 1 are unaltered by a specification of values of local beables in a space-like separated region 2 when what happens in the backward light cone is already *sufficiently* specified, for example by a full specification of local beables in a space-time region 3. It is important that region 3 completely shields off from 1 the overlap of the backward light cones

of 1 and 2. And it is important that events in 3 be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. *The hypothesis is that any such information about 2 becomes redundant when 3 is specified completely.* J.S. Bell 2004 p. 240.

and

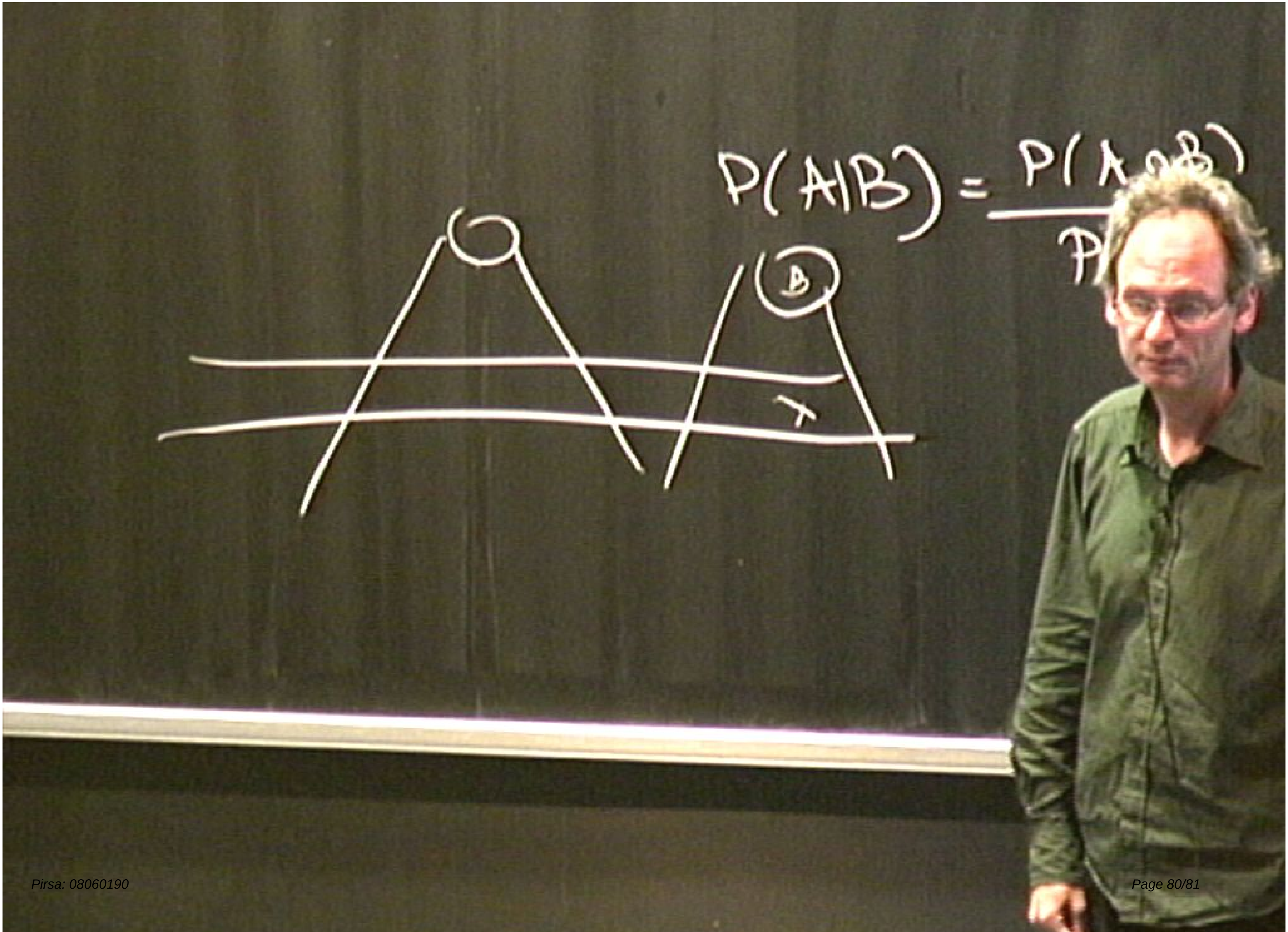
"Invoking local causality AND THE AS-of 1 and 2. And it is important that events in 3 be specified completely. Otherwise the traces in region 2 of causes of events in 1 could well supplement whatever else was being used for calculating probabilities about 1. *The hypothesis is that any such information about 2 becomes redundant when 3 is specified completely.* J.S. Bell 2004 p. 240.

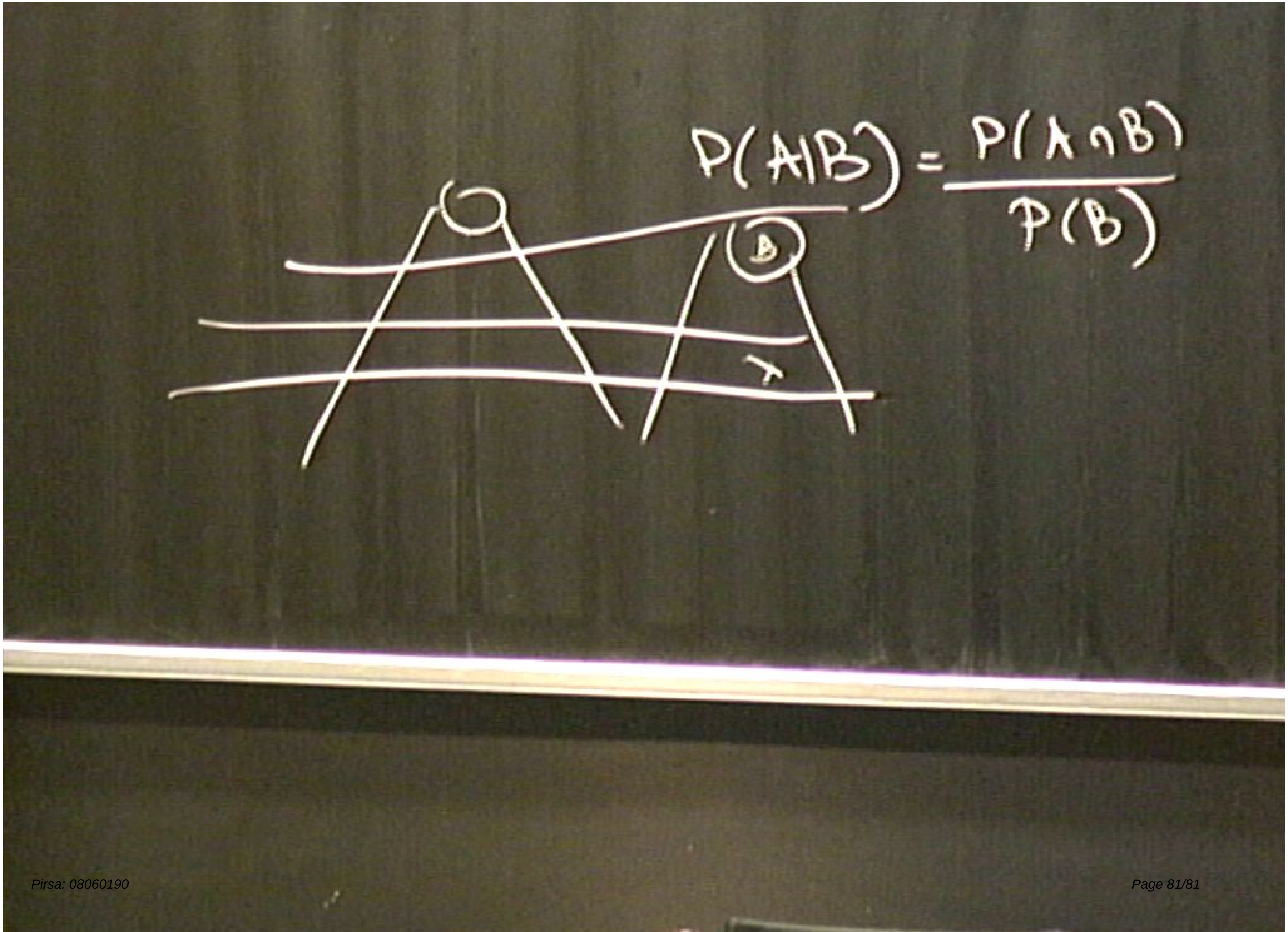
and

"Invoking local causality AND THE ASSUMED COMPLETENESS OF c and λ,. . . we have

$$p_{\alpha\beta}(a,b|\lambda c) = p_{\alpha}(a|\lambda c)p_{\beta}(b|\lambda c)"$$

(ibid, p. 243. Emphasis added)


P(A|B) =
B


$P(A|B) = \frac{P(A \cap B)}{P(B)}$