Title: Newcomb\'s problem and Bell\'s theorem

Date: Sep 30, 2008 04:30 PM

URL: http://pirsa.org/08090086

Abstract: In recent years there has been a growing awareness that studies on quantum foundations have close relationships with other fields such as probability and information theory. In this talk I give another example of how such interdisciplinary work can be fruitful, by applying some of the lessons from quantum mechanics, in particular from Bell\'s theorem, to a debate on the philosophical foundations of decision theory. I argue that the basic assumptions of the popular causal decision theory -- which was developed partly in response to a puzzle proposed by the physicist William Newcomb and published by the philosopher Robert Nozick -- are analogous to the basic assumptions of a local hidden-variables theory in the context of Bell\'s theorem. Both have too strong a prejudice about the causal structure of the world: there are possible games the world can pose such that an agent who operates by those theories is constrained to choose losing strategies no matter what evidence he or she acquires.

Griffith Centre for Quantum Dynamics

Griffith UNIVERSITY

THE UNIVERSITY OF QUEENSLAND

The University of Sydney

Newcomb's problem

Newcomb's problem

Newcomb's problem

Newcomb's problem

- You can pick

  A_1) just Box 1

  A_2) both

Eric Cavalcanti, Waterloo, October 2008

Newcomb's problem

- You can pick

  A_1) just Box 1

  A_2) both

- Predictor already put $1 Million in the box if and only if she predicted you were going to choose just Box 1

First solution: classical Bayesian (Evidential) Decision Theory

- Maximise expected conditional utility:

$$EEU(A) \;=\; \sum_{O} P(O|A)\,u(O,A)$$

- In the Newcomb scenario:

$$EEU(A_1) = P(M|A_1)\,10^6$$
$$EEU(A_2) = P(M|A_2)\,10^6 + 10^3$$

- But $P(M|A_1) \gg P(M|A_2)$

$\Longrightarrow$ Take one box only

On the other hand: principle of dominance

- No matter what's in box 1, I'm better off taking both

- Critique: dominance isn't an independent principle, but only a short-cut, and can only be used in case of no probabilistic dependence from actions to outcomes.

- Response: dominance should be used when the choices cannot *causally influence* the outcomes.

Nozick: paradox of rationality

- For Robert Nozick, Newcomb's problem displayed a paradox for rationality. Both solutions have equal support; intuitions vary with the contents of the boxes.

Causal decision theory (CDT)

- Intuition behind the dominance argument is formalised in *Causal Decision Theory*.
- Conditional probabilities in expected utility should be *causal* (or counterfactual, under a causal reading of the counterfactual)

$$CEU(A) \;=\; \sum_O P_C(O|A)\, u(O,A)$$

- In the Newcomb problem in particular:

$$P_C(M|A_1) = P_C(M|A_2)$$

$\Rightarrow$ Take both boxes

Causal decision theory (CDT)

Causal probabilities can be understood as an unconditional average over "dependency hypotheses" (Lewis) or "causal propensities" (Skyrms). Denoted by a set of variables 'K'.

$$P_C(O|A) = \sum_K P(K)P(O|A;K)$$

As opposed to the actual conditional probability

$$\begin{aligned} P(O|A) &= \sum_K P(K,O|A) \\ &= \sum_K P(K|A)P(O|A;K) \end{aligned}$$

“Effective” probabilities

Define “effective” probability as whatever you should use in your decisions.

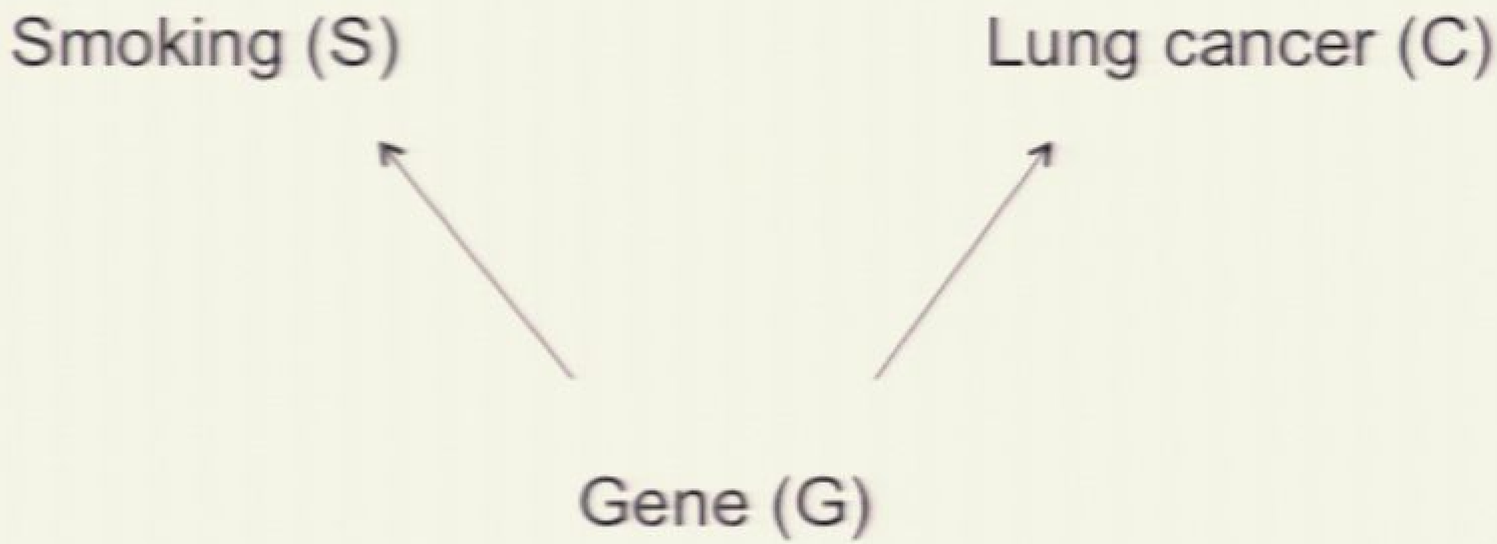
For CDT these can be different in general:

- The usual, evidential, conditional probabilities.
- The effective probabilities. For CDTists, causal probabilities.

(What does the first mean if it doesn’t have any practical influence?)

A Newcomb-type problem can be posed whenever they differ.

“Medical” Newcomb problems



“Medical” Newcomb problems

"Medical" Newcomb problems

Genetics and lung cancer

Smoking out the smoking gene

Apr 3rd 2008 | NEW YORK

From *The Economist* print edition

Your genes may control how much you smoke—and how likely you are to get lung cancer as a result

Alamy

The "tickle" defence of EDT (Horgan)

- The action of the gene could only be through the agent's beliefs and desires. She would feel a "tickle" T that tells her that she desires smoking.

- Conditional on T, cancer is screened off from the choice of smoking.

$$P(C|S;T) = P(C|\neg S;T) = P(C|T)$$

$$P(C|S;\neg T) = P(C|\neg S;\neg T) = P(C|\neg T)$$

- A rational agent should take into account *all* available evidence

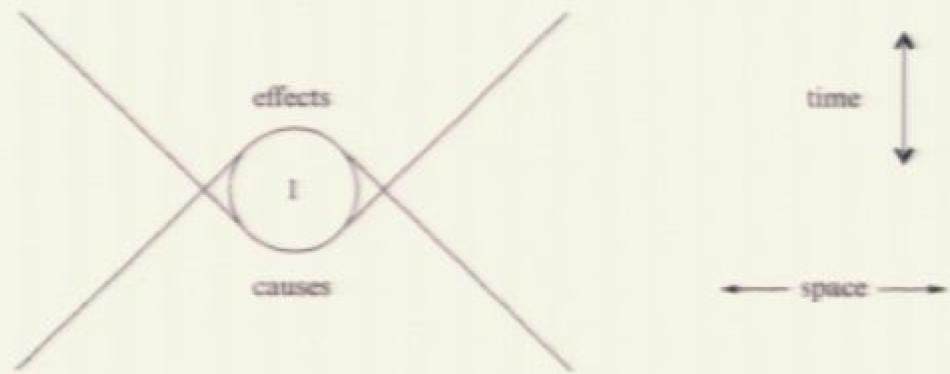
$\Rightarrow$ Bayesian decision theory advises smoking as well.

Is CDT not even wrong?

- Horgan and others (e.g. Eells, Price) essentially argue that causal probabilities and evidential probabilities don't actually come apart in problems such as the smoking gene.
- CDT – just an irrelevant complication?
- But what if in some cases they *do* come apart?
- Those defences are not general enough.

Regions of causal influence

- CDT needs an account of what's "inside" and "outside" the causal influence of an agent.
- Depends on the agent's causal theory, or the "dependency hypotheses" *K*.
- Relativity:
  - Inside: future light cone
  - Outside: everywhere else



Regions of causal influence

b: any factors *outside* the agents' causal influence

$$P(b|A;K) = P(b|A\ ;K) = P(b|K)$$

a: any factors *inside* the agents' causal influence

$$P(a,b|A;K) = P(a|A;K)P(b|K)$$

Therefore when a decision situation depends on *a* **and** *b* CDT should use

$$P_C(a,b|A) = \sum_K P(K)P(a|A;K)P(b|K)$$

The parallel with Bell's theorem

"Free will" / free conditionalisation / no-retrocausality

$$P(\lambda|A,B) = P(\lambda)$$

Local causality

$$P(a,b|A,B;\lambda) = P(a|A;\lambda)P(b|B;\lambda)$$

Therefore

$$P_{LHV}(a,b|A,B) = \sum_K P(\lambda)P(a|A;\lambda)P(b|B;\lambda)$$

The parallel with Bell's theorem

Compare

$$P_C(a,b|A) = \sum_K P(K)P(a|A;K)P(b|K)$$

$$P_{LHV}(a,b|A,B) = \sum_K P(\lambda)P(a|A;\lambda)P(b|B;\lambda)$$

Eric Cavalcanti, Waterloo, October 2008

Bell (1964)

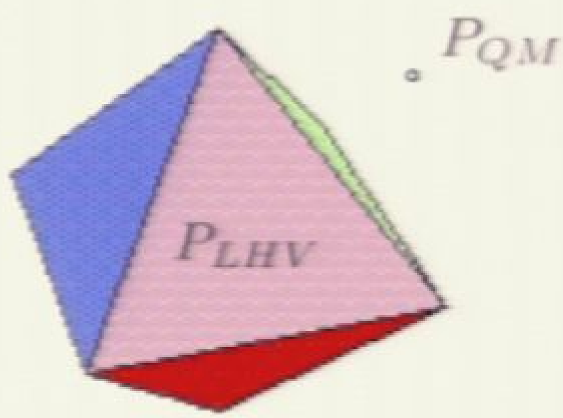
No Local Hidden Variable model like that can explain the correlations predicted by Quantum Mechanics between certain entangled pairs of particles

Aspect (1980's)

And what's more, we've tested the predictions that
Bell worked out, and found
that quantum theory *is* accurate.

The causal polytope

- In Bell inequalities, it is common to analyse the set of allowed LHV probabilities
- These live in a convex polytope

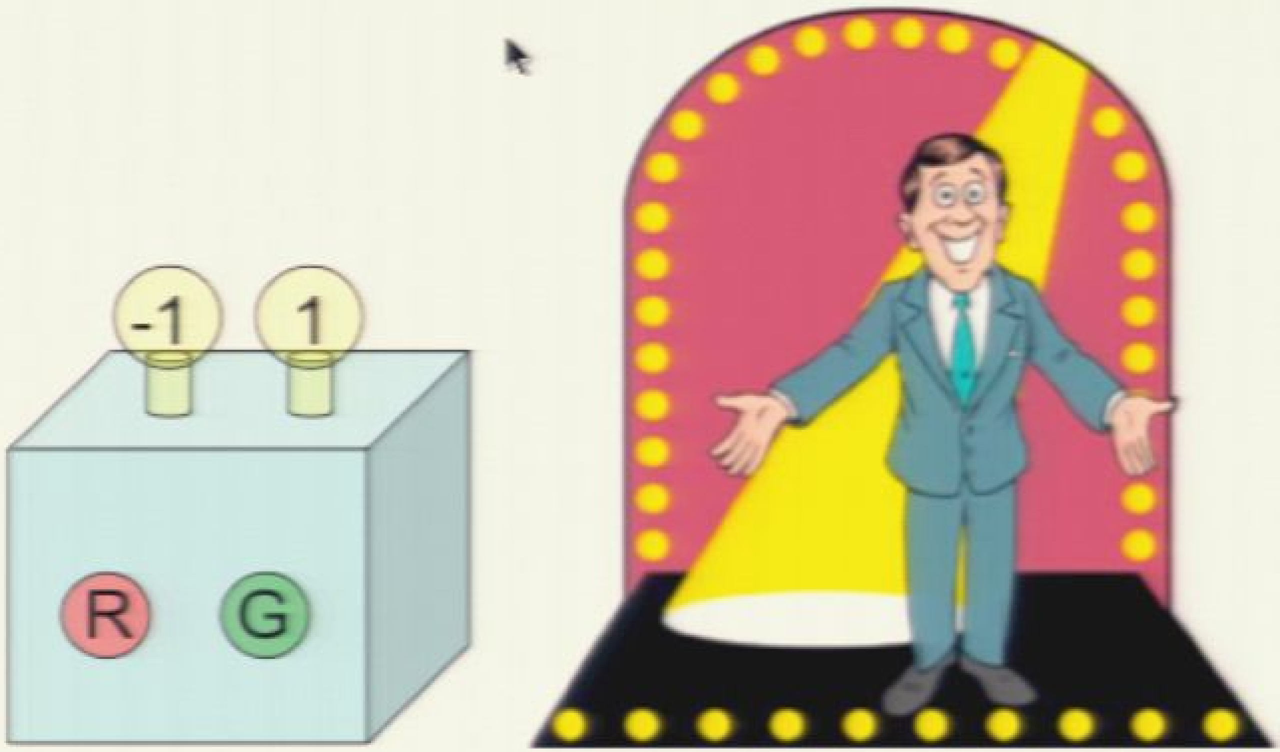


- Quantum probabilities can lie outside the causal polytope

- A CDTist will be constrained to have his causal probabilities in the polytope no matter what evidence is thrown at them

The Bell game

The Bell game



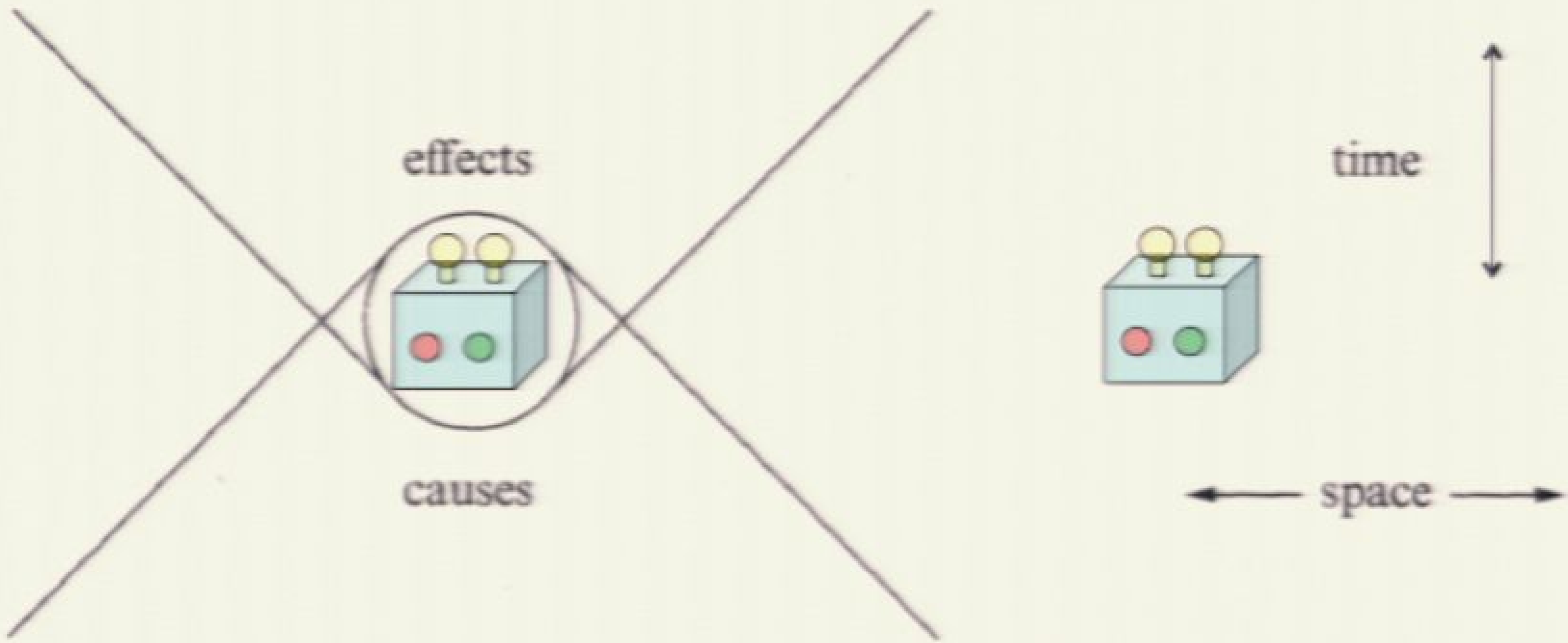
Alice

The Bell game

Alice

Bob

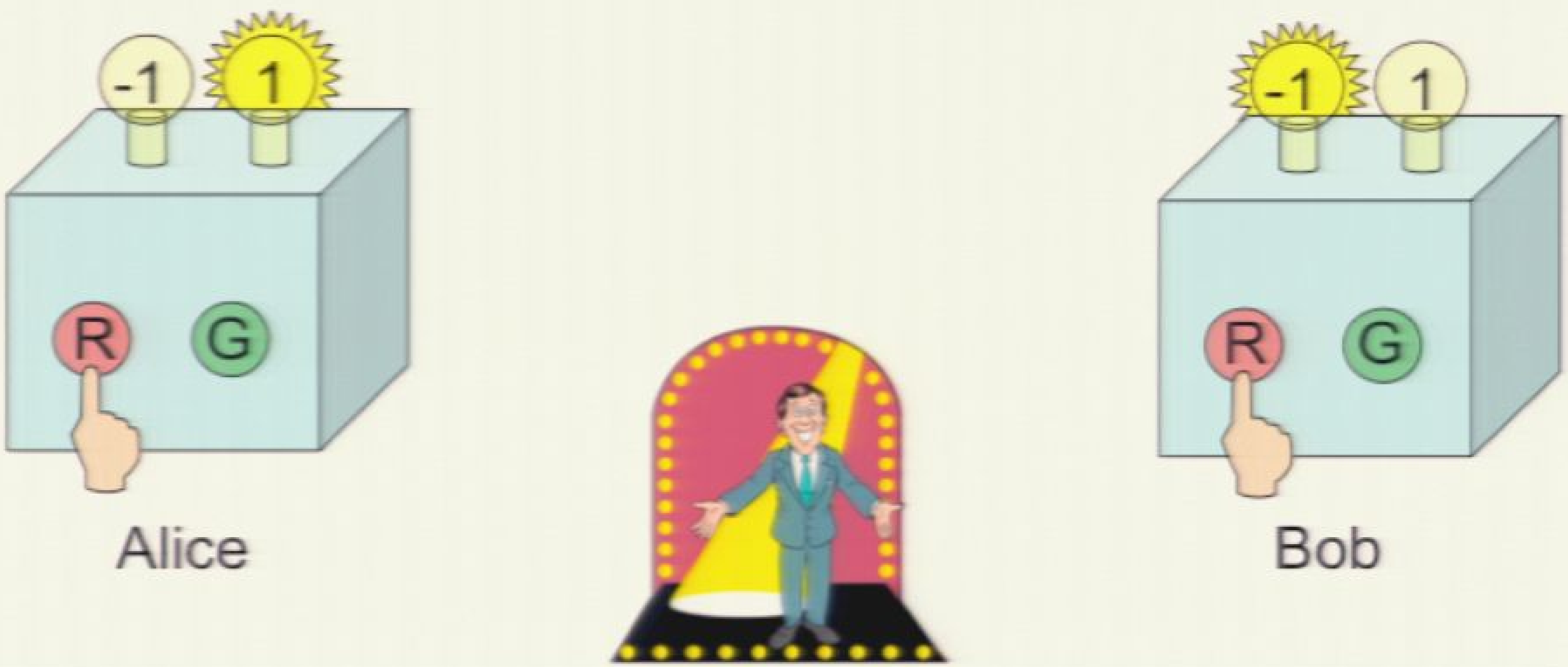
The Bell game



The Bell game

The Bell game

Run 1: $R_A R_B = -1$



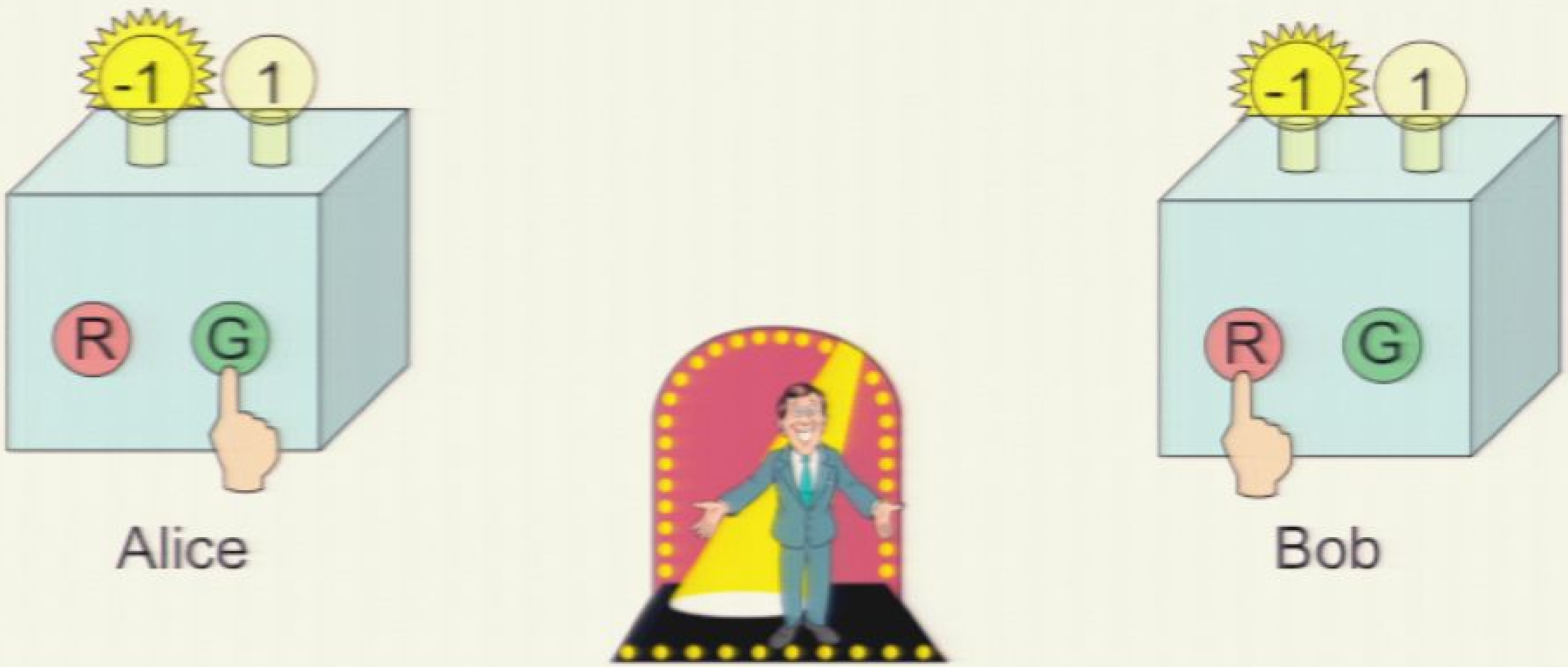
The Bell game

Run 1: $R_A R_B = -1$

The Bell game

Run 1: $R_A R_B = -1$

Run 2: $G_A R_B = 1$



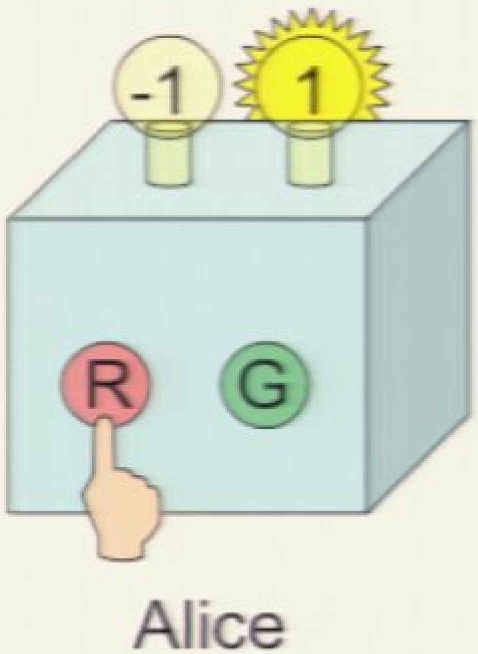
Eric Cavalcanti, Waterloo, October 2008

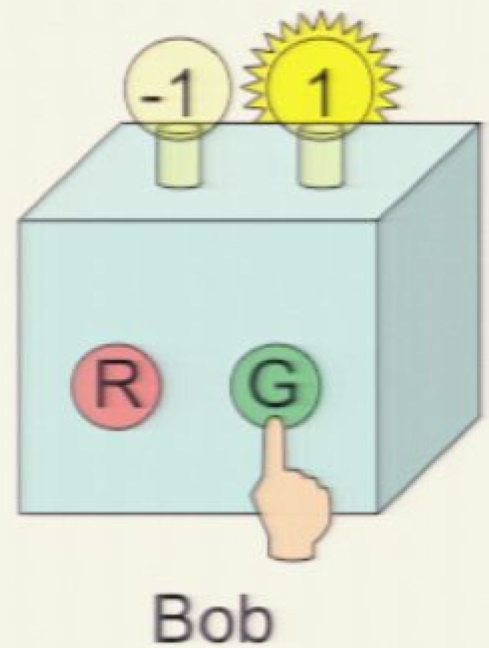
The Bell game

Run 1: $R_A R_B = -1$

Run 2: $G_A R_B = 1$

Run 3: $R_A G_B = 1$





The Bell game

Run 1: $R_A R_B = -1$

Run 2: $G_A R_B = 1$

Run 3: $R_A G_B = 1$

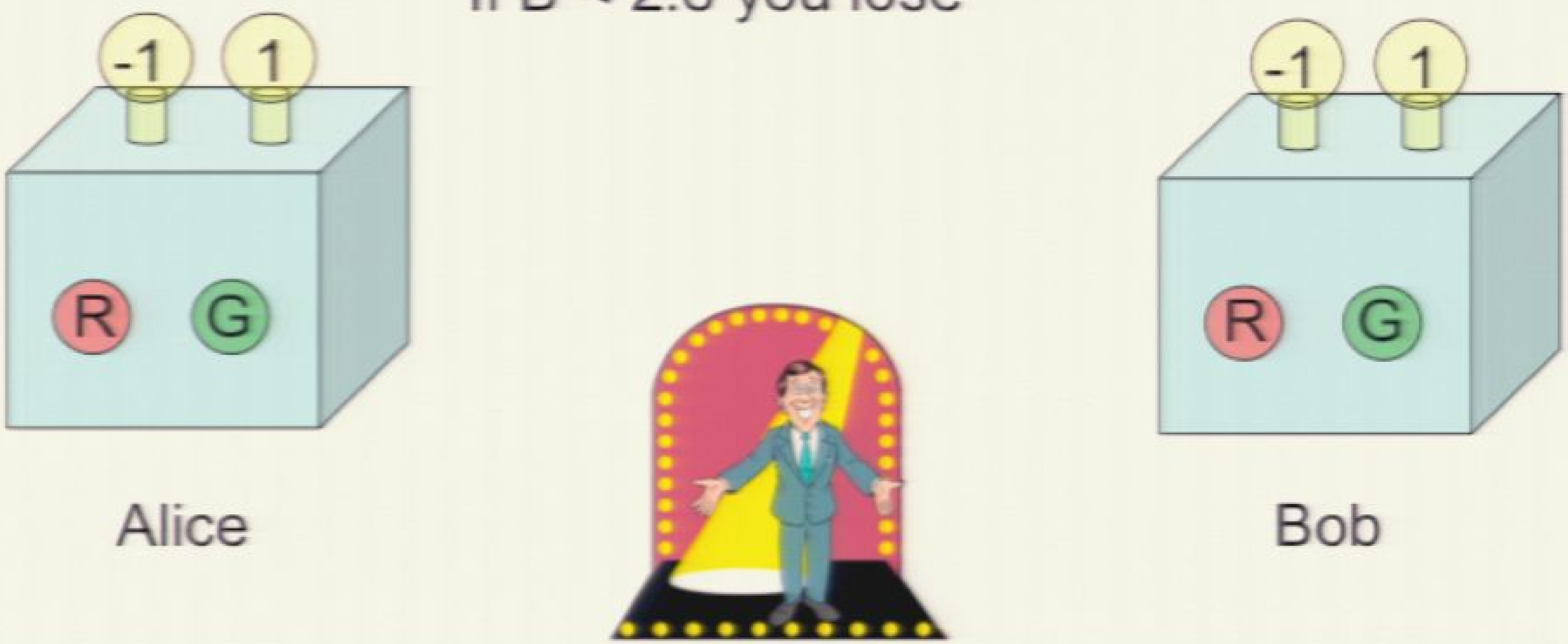
...

The Bell game

$$B = \langle R_A R_B \rangle + \langle G_A R_B \rangle + \langle R_A G_B \rangle - \langle G_A G_B \rangle$$

If B >= 2.8 you win $1 M

If B < 2.8 you lose



The Bell game

$$B = \langle R_A R_B \rangle + \langle G_A R_B \rangle + \langle R_A G_B \rangle - \langle G_A G_B \rangle$$

If B >= 2.8 you win $1 M

If B < 2.8 you lose


$ 1 K

Or you can just take home $1000…

Analysis of the game

- CDT should use the causal probabilities

$$P_C(a,b|A,B) = \sum_K P(K) P(a|A;K) P(b|B;K)$$

- For each pair of Alice's and Bob's choices

$$\begin{aligned}\langle AB \rangle &= \sum_K P(K) \sum_{a,b} a\, b\, P(a|A;K) P(b|B;K) \\ &= \sum_K P(K) \sum_a a\, P(a|A;K) \sum_b b\, P(b|B;K) \\ &= \sum_K P(K) \langle A \rangle_K \langle B \rangle_K\end{aligned}$$

Analysis of the game

- Bob's formula becomes

$$\begin{aligned} B &= \langle R_A R_B \rangle + \langle R_A G_B \rangle + \langle G_A R_B \rangle - \langle G_A G_B \rangle \\ &= \sum_K P(K) \{ \langle R_A \rangle_K (\langle R_B \rangle_K + \langle G_B \rangle_K) + \langle G_A \rangle_K (\langle R_B \rangle_K - \langle G_B \rangle_K) \} \\ &\leq \sum_K P(K)\, 2 = 2 \end{aligned}$$

- QM predicts

$$B = 2\sqrt{2} \approx 2.83$$

Play: B >= 2.8 you win $1 M
B < 2.8 you lose
Or take the risk-free $1000

CDT says take the thousand dollars;
EDT says play the game

Analysis of the game

Lewis: “They have their millions and we have our thousands. They think this goes to show the error of our ways, [but] we have no choice. The riches are reserved for the irrational”.

Possible objections (1)

- “The game is not exactly analogous to the original Newcomb problem”.

- No, but CDT should be applied in every decision situation. The important thing for a Newcomb-type problem is a disagreement between the evidential and causal probabilities.

- The advantage of this game is that you can actually do it.

Possible objections (2)

"The money in the box is *actually known by someone*. But the hidden variables in QM are hidden even in principle. Suppose there's a friend of mine looking at the closed box. Surely he would advise me to take both boxes?"

Reply:

- Nothing in causal decision theory requires the posited causal factors to be known by someone (e.g., the gene).
- Your friend would advise you to take both boxes *no matter what*. He gives you no information.
- An agent's choices can only depend on *their* information, not on someone else's.
- The objection reveals a fragility of the original Newcomb scenarios, not of my argument…

Communicated vs. non-communicated predictions

- What makes the "tickle defences" work is that they argue that those are effectively cases of communicated prediction.

- However, the Predictor can *know* the effect of the communication on the prediction itself. She can't always communicate a prediction and still keep it accurate!

- A way to guarantee that the prediction won't be falsified is to guarantee that the knowledge simply won't be available to the agent. HVs are inaccesible even in principle.

Possible objections (3)

"The argument seems to depend on a 'superdeterministic' interpretation of the Bell correlations. What if I believe in non-local causality?".

- Reply: there is no generally accepted way of explaining quantum correlations as causal correlations. Our best theory of causal structure is relativity. Why not get your causal probabilities from relativity? (It can't be because it doesn't give the right *evidential* probabilities!)

- Superdeterminism is a *logically possible* explanation, and it agrees with relativity as far as local causality is concerned. If you give this hypothesis any nonzero credibility, the argument holds.

Possible objections (4)

- "Alice *does* causally influence the correlations. It is her choice of a biased ensemble which *causes* the Bell violation to occur"

  - "Extended" Newcomb problem:

    Instead of one closed box, there are 100.

    Alice can choose to

    (a) take all of them and the extra thousand; or

    (b) open just one.

    As before, P(M|b) >> P(M|a)

  - Is it plausible to argue that it is Alice's choice of *which* of the hundred boxes to take that *causes* the money to be there?

Lessons *for* physics?

- Possible defences from CDT camp: new loopholes in Bell's theorem?
- More attention to the "no-retrocausality" / "free-will" assumption – e.g., retrocausal models of QM. Advantage: saves local causality.
- Introduce explicitly the agent's choices of experiments in searches for information-theoretic principles for QM.