The Great Wall of China Five Point Plan for Unification One World One Dream

### One Boxer Rebellion

Huw Price

9 January 2009

- The Great Divide
- 2 The First Principle of Causal Epistemology
- Working From the Inside
- Sticking to One's Guns
- The Problem of the 'Medical' Cases

## James Joyce on the Great Wall of China

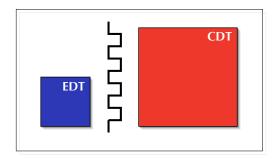
# James Joyce on the Great Wall of China



# One World One Dream



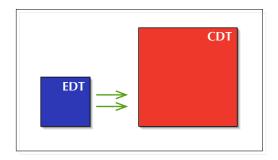
# The orthodox picture



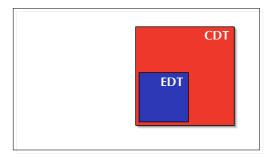
### 1. Tear down the wall



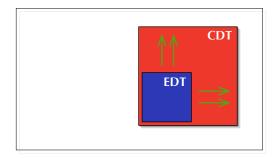
### 2. EDTers should be CDTers, too



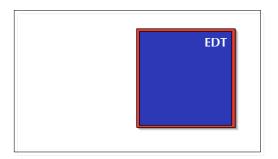
### 3. We are all CDTers now



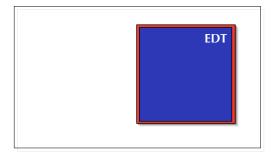
# 4. Causation is evidential



# 5. One-boxing wins



### One World One Dream One Box



- The Great Divide
- The First Principle of Causal Epistemology
- Working From the Inside
- Sticking to One's Guns
- The Problem of the 'Medical' Cases

### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B
- We are thereby committed to view that A is a (positive) causal factories for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations as

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And *vice versa?*)
- Nicely illustrated by Eric's game, based on the Bell correlations

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations.

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations.

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations.

#### The First Principle

"In the game of discerning causal links, effective strategies are trumps."

- Suppose we take it to be rational to do A rather than not-A, in pursuit
  of an end B i.e., we take A to be an effective strategy for achieving B.
- We are thereby committed to view that A is a (positive) causal factor for B.
- (And vice versa?)
- Nicely illustrated by Eric's game, based on the Bell correlations.

- The Great Divide
- 2 The First Principle of Causal Epistemology
- Working From the Inside
- Sticking to One's Guns
- The Problem of the 'Medical' Cases

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First Principle ensures that my causal judgements track my decisions about effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer respond:

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First
  Principle ensures that my causal judgements track my decisions about
  effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer respond?

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First Principle ensures that my causal judgements track my decisions about effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer respond?

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First Principle ensures that my causal judgements track my decisions about effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer respond:

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First Principle ensures that my causal judgements track my decisions about effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer respond

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be evidential.
- For me, then, there's no gap between CDT and EDT the First Principle ensures that my causal judgements track my decisions about effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

How can a two-boxer responds

#### Consider a CDTer who says:

- I accept the First Principle (of course).
- I take strategic deliberation to be *evidential*.
- For me, then, there's no gap between CDT and EDT the First
  Principle ensures that my causal judgements track my decisions about
  effective strategies, as determined by EDT.
- In the (classic) Newcomb Problem I therefore one-box, taking my action to affect the Predictor's choice.

#### How can a two-boxer respond?

### "You're wrong about the causal structure of the Newcomb case."

Reply: The First Principle ensures that we can't know this, without already having decided the issue about effective strategies — while that issue is *open*, causal structure cannot be regarded as however.

"You're wrong about the causal structure of the Newcomb case."

REPLY:

The First Principle ensures that we can't know this, without already having decided the issue about effective strategies – while that issue is *open*, causal structure cannot be regarded as *known*.

"You're wrong about the causal structure of the Newcomb case."

Reply: The First Principle ensures that we can't know this, without already having decided the issue about effective strategies – while that issue is *open*, causal structure cannot be regarded

"You're wrong about the causal structure of the Newcomb case."

Reply: The First Principle ensures that we can't know this, without already having decided the issue about effective strategies – while that issue is *open*, causal structure cannot be regarded as *known*.

### "Let's just stipulate the causal structure."

REPLY:

This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has no coherent solution.<sup>2</sup>)

<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

### "Let's just stipulate the causal structure." 1

REPLY:

This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has

no coherent solution.<sup>2</sup>)

<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

### "Let's just stipulate the causal structure." 1

REPLY:

This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has no coherent solution.)

no coherent solution.<sup>2</sup>)

<sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem — more on this later



<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions

### "Let's just stipulate the causal structure." 1

Reply: This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has

<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions.

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

### "Let's just stipulate the causal structure." 1

Reply: This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal

structure. (It isn't a surprise that an incoherent problem has

no coherent solution.<sup>2</sup>)

<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions.

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

### "Let's just stipulate the causal structure." 1

REPLY: This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has no coherent solution.

<sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

### "Let's just stipulate the causal structure." 1

Reply: This makes the description of the case self-contradictory, by my agent's lights. You've specified a causal structure, and provided information that implies a different causal structure. (It isn't a surprise that an incoherent problem has no coherent solution.<sup>2</sup>)

<sup>&</sup>lt;sup>1</sup>In other words, let's set up the case with the stipulation that the agent cannot affect the Predictor's actions.

<sup>&</sup>lt;sup>2</sup>Though there's a nice explanation here for the apparent intractability of the Newcomb Problem – more on this later.

## "If you'd taken two boxes, you would have been richer."

REPLY:

We don't know the counterfactuals until we know the causal structure, and that's precisely what's at issue. By my agent's lights, it simply isn't true that he would have been richer if he'd two-boxed – on the contrary!

## "If you'd taken two boxes, you would have been richer."

Reply:

We don't know the counterfactuals until we know the causal structure, and that's precisely what's at issue. By my agent's lights, it simply isn't true that he would have been richer if he'd two-boxed – on the contrary!

"If you'd taken two boxes, you would have been richer."

REPLY: We don't know the counterfactuals until we know the causal structure, and that's precisely what's at issue. By my agent's lights, it simply isn't true that he would have been richer if he'd two-boxed – on the contrary!

"If you'd taken two boxes, you would have been richer."

Reply: We don't know the counterfactuals until we know the causal structure, and that's precisely what's at issue. By my agent's lights, it simply isn't true that he would have been richer if he'd two-boxed – on the contrary!

- The Newcomb puzzle stems from failing to get back to the First Principle failing to see the *priority* of a notion of effective strategy with respect to our notion of causation.
- This failure encourages us to regard causation as an independent degree of ontological freedom, so that we think we can imagine Newcomb-like cases in which CDT and EDT come apart.
- But an EDTer who keeps her head, rejects that kind of decisionindependent causal ontology, and insists that she is a CDTer too, can resist this move.
- Moreover, she then has a trump card . . .

- The Newcomb puzzle stems from failing to get back to the First Principle failing to see the *priority* of a notion of effective strategy, with respect to our notion of causation.
- This failure encourages us to regard causation as an independent degree of ontological freedom, so that we think we can imagine Newcomb-like cases in which CDT and EDT come apart.
- But an EDTer who keeps her head, rejects that kind of decisionindependent causal ontology, and insists that she is a CDTer too, can resist this move.
- Moreover, she then has a trump card . . .

- The Newcomb puzzle stems from failing to get back to the First Principle failing to see the *priority* of a notion of effective strategy, with respect to our notion of causation.
- This failure encourages us to regard causation as an independent degree of ontological freedom, so that we think we can imagine Newcomb-like cases in which CDT and EDT come apart.
- But an EDTer who keeps her head, rejects that kind of decisionindependent causal ontology, and insists that she is a CDTer too, can resist this move.
- Moreover, she then has a trump card . . .

- The Newcomb puzzle stems from failing to get back to the First Principle failing to see the *priority* of a notion of effective strategy, with respect to our notion of causation.
- This failure encourages us to regard causation as an independent degree of ontological freedom, so that we think we can imagine Newcomb-like cases in which CDT and EDT come apart.
- But an EDTer who keeps her head, rejects that kind of decisionindependent causal ontology, and insists that she is a CDTer too, can resist this move.
- Moreover, she then has a trump card . . .

- The Newcomb puzzle stems from failing to get back to the First Principle failing to see the *priority* of a notion of effective strategy, with respect to our notion of causation.
- This failure encourages us to regard causation as an independent degree of ontological freedom, so that we think we can imagine Newcomb-like cases in which CDT and EDT come apart.
- But an EDTer who keeps her head, rejects that kind of decisionindependent causal ontology, and insists that she is a CDTer too, can resist this move.
- Moreover, she then has a trump card . . .

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways.

The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational When we made our choices, there were no millions to be had. In the words of Gibbard and Harper.

we take the moral... to be something else: if someone is very good a predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not:

#### • Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways.

#### The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational When we made our choices, there were no millions to be had. In the words of Gibbard and Harper.

we take the moral ... to be something else: if someone is very good an predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways.

The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

 But this response is now blocked – we now have two different conceptions of causal structure, and hence of rational CDT-guided choice – one leading

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper.

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.



• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral . . . to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.



• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral . . . to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

• Lewis on a favourite one-boxer argument:

The one-boxers sometimes taunt us: if you're so smart, why ain'cha rich? They have their millions and we have our thousands, and they think this goes to show the error of our ways. [Indeed! – HP]

• The two-boxer response:

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral . . . to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
  - If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

- Newcomb's Problem turns on a clash between two conceptions of causality:
  - A pragmatic notion, with its roots in the soil of practical (evidential) decision making.
  - A notion corrupted by foreign metaphysical influences by theories of causation which have lost sight of their practical origins.
- If we're corrupted by foreign influences, the local methods may well seem irrational.
- But the fault lies with the foreign causal principles, which are unsuited to local conditions, in the strange world of Newcomb's Problem.

# Options for two-boxers

#### Two options at this point for my opponents:

- Try to "own" the notion of effective strategy, so that the First Principle delivers two-boxing.
- Appeal to more realistic Newcomb puzzles e.g., the so-called "medical" cases – in which it seems far less plausible to claim that the true causal structure is non-standard.

# Options for two-boxers

#### Two options at this point for my opponents:

- Try to "own" the notion of effective strategy, so that the First Principle delivers two-boxing.
- Appeal to more realistic Newcomb puzzles e.g., the so-called "medical" cases – in which it seems far less plausible to claim that the true causal structure is non-standard.

# Options for two-boxers

#### Two options at this point for my opponents:

- Try to "own" the notion of effective strategy, so that the First Principle delivers two-boxing.
- Appeal to more realistic Newcomb puzzles e.g., the so-called "medical" cases in which it seems far less plausible to claim that the true causal structure is non-standard.

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

- The intended reading here depends on the counterfactual, "If we had two-boxed, we would have been (even) richer," and we've already responded to that: assuming the counterfactual is the same as assuming the disputed causal structure.
- So let's try an unintended reading . . . with support in the Great Texts

#### • Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

- The intended reading here depends on the counterfactual, "If we had
  two-boxed, we would have been (even) richer," and we've already responded to
  that: assuming the counterfactual is the same as assuming the disputed causal
  structure.
- So let's try an unintended reading . . . with support in the Great Texts.

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

- The intended reading here depends on the counterfactual, "If we had two-boxed, we would have been (even) richer," and we've already responded to that: assuming the counterfactual is the same as assuming the disputed causal structure.
- So let's try an unintended reading . . . with support in the Great Texts.

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

- The intended reading here depends on the counterfactual, "If we had
  two-boxed, we would have been (even) richer," and we've already responded to
  that: assuming the counterfactual is the same as assuming the disputed causal
  structure.
- So let's try an unintended reading ... with support in the Great Texts

# "Why ain'cha rich?," again

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

- The intended reading here depends on the counterfactual, "If we had
  two-boxed, we would have been (even) richer," and we've already responded to
  that: assuming the counterfactual is the same as assuming the disputed causal
  structure.
- So let's try an unintended reading ... with support in the Great Texts.

# "Why ain'cha rich?," again

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

- The intended reading here depends on the counterfactual, "If we had two-boxed, we would have been (even) richer," and we've already responded to that: assuming the counterfactual is the same as assuming the disputed causal structure.
- So let's try an unintended reading ... with support in the Great Texts.

# "Why ain'cha rich?," again

• Lewis's response to "Why ain'cha rich?":

The reason why we are not rich is that the riches were reserved for the irrational. When we made our choices, there were no millions to be had. In the words of Gibbard and Harper,

we take the moral ... to be something else: if someone is very good at predicting behavior and rewards predicted irrationality richly, then irrationality will be richly rewarded.

Rationality will not.

- The intended reading here depends on the counterfactual, "If we had two-boxed, we would have been (even) richer," and we've already responded to that: assuming the counterfactual is the same as assuming the disputed causal structure.
- So let's try an unintended reading ... with support in the Great Texts.

For when the One Great Scorer comes To write against your name, He marks – not that you won or lost – But how you played the game.

For when the One Great Scorer comes To write against your name, He marks – not that you won or lost – But how you played the game.

- Grantland Rice, "Alumnus Football."





For what doth it profit a man, to gain the whole world, and forfeit his [decision principles]?

- Matthew, 8:36



- Sure, one-boxing makes you "rich".
- But there are loftier goals than money.
- Which would you rather be, irrational and "rich", or rational but not "rich"?

- Sure, one-boxing makes you "rich".
- But there are loftier goals than money!
- Which would you rather be, irrational and "rich", or rational but not "rich"?



- Sure, one-boxing makes you "rich".
- But there are loftier goals than money!
- Which would you rather be, irrational and "rich", or rational but not "rich"?



- Sure, one-boxing makes you "rich".
- But there are loftier goals than money!
- Which would you rather be, **irrational** and "rich", or **rational** but not "rich"?



- Someone who is playing with the One Great Scorer's opinion in mind is still playing to win – it's just that the game isn't football any more!
- Similarly, it isn't incoherent to assign utilities to strategies, but that just changes the decision problem.
- What we are interested in is simply what constitutes an effective utility-enhancing strategy for someone with the original utilities (represented for convenience by money).

- Someone who is playing with the One Great Scorer's opinion in mind is still playing to win – it's just that the game isn't football any more!
- Similarly, it isn't incoherent to assign utilities to strategies, but that just changes the decision problem.
- What we are interested in is simply what constitutes an effective utility-enhancing strategy for someone with the original utilities (represented for convenience by money).

- Someone who is playing with the One Great Scorer's opinion in mind is still playing to win – it's just that the game isn't football any more!
- Similarly, it isn't incoherent to assign utilities to strategies, but that just changes the decision problem.
- What we are interested in is simply what constitutes an effective utility-enhancing strategy for someone with the original utilities (represented for convenience by money).



- Someone who is playing with the One Great Scorer's opinion in mind is still playing to win – it's just that the game isn't football any more!
- Similarly, it isn't incoherent to assign utilities to strategies, but that just changes the decision problem.
- What we are interested in is simply what constitutes an effective utility-enhancing strategy for someone with the original utilities (represented for convenience by money).



- Someone who is playing with the One Great Scorer's opinion in mind is still playing to win – it's just that the game isn't football any more!
- Similarly, it isn't incoherent to assign utilities to strategies, but that just changes the decision problem.
- What we are interested in is simply what constitutes an effective utility-enhancing strategy, for someone with the original utilities (represented for convenience by money).



- How do we discover effective strategies
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (Cf. "Interventionism")
- Examples of such effective-strategy-supporting correlations
  - The Bell correlations
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (Cf. "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations.
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations.
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling *A*, and look for correlations with the movements of *B*.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations.
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations.
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

- How do we discover effective strategies?
- We "try things and see" we try wiggling A, and look for correlations with the movements of B.
- In other words, we look for correlations which survive "free choice" of the putative causal variable. (*Cf.* "Interventionism")
- Examples of such effective-strategy-supporting correlations:
  - The Bell correlations.
  - (By assumption) the correlation between one-boxing and wealth, in the standard Newcomb problem.

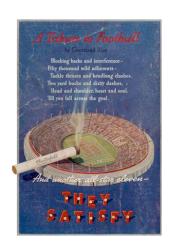
- The Great Divide
- 2 The First Principle of Causal Epistemology
- Working From the Inside
- Sticking to One's Guns
- The Problem of the 'Medical' Cases

- Imagine there's a gene that predisposes both to smoking and cancer ...
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will general cancer?
- Isn't EDT committed to saying "Yes!"?

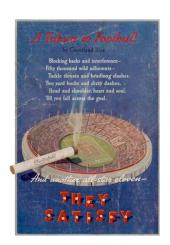
- Imagine there's a gene that predisposes both to smoking and cancer ...
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will get cancer?
- Isn't EDT committed to saving "Yes!"?

- Imagine there's a gene that predisposes both to smoking and cancer ...
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will get cancer?
- Isn't EDT committed to saying "Yes!"?

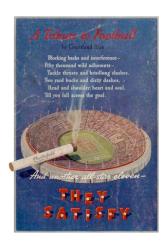
- Imagine there's a gene that predisposes both to smoking and cancer ...
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will get cancer?
- Isn't EDT committed to saying "Yes!"?



- Imagine there's a gene that predisposes both to smoking and cancer . . .
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will get cancer?
- Isn't EDT committed to saying "Yes!"?



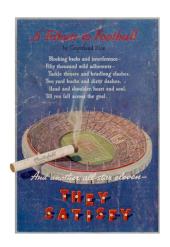
- Imagine there's a gene that predisposes both to smoking and cancer . . .
- ... and that smoking predisposes to satisfaction!
- Should you refrain from smoking, to reduce the probability that you will get cancer?
- Isn't EDT committed to saying "Yes!"?



### Causal information is evidential information

### My response

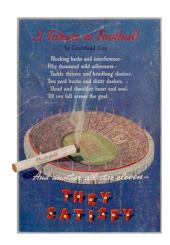
- When we're told the causal structure of the case (in an ordinary case like this!) we're told the relevant evidential dependencies – after all, that's what causation is!
- The puzzle is just that of explaining why the statistical correlation (of smoking with possession of the gene) doesn't translate into an evidential dependency, from the agent's point of view.
- That's the puzzle the "Tickle Defence" and its descendants address.



### Causal information is evidential information

### My response

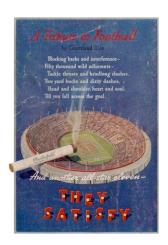
- When we're told the causal structure of the case (in an ordinary case like this!) we're told the relevant evidential dependencies – after all, that's what causation is!
- The puzzle is just that of explaining why the statistical correlation (of smoking with possession of the gene) doesn't translate into an evidential dependency, from the agent's point of view.
- That's the puzzle the "Tickle Defence" and its descendants address.



### Causal information is evidential information

### My response

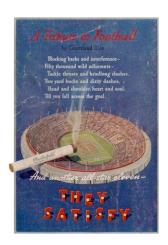
- When we're told the causal structure of the case (in an ordinary case like this!) we're told the relevant evidential dependencies – after all, that's what causation is!
- The puzzle is just that of explaining why the statistical correlation (of smoking with possession of the gene) doesn't translate into an evidential dependency, from the agent's point of view.
- That's the puzzle the "Tickle Defence" and its descendants address.



### Causal information is evidential information

### My response

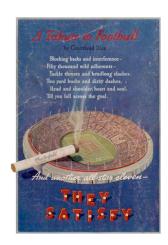
- When we're told the causal structure of the case (in an ordinary case like this!) we're told the relevant evidential dependencies – after all, that's what causation is!
- The puzzle is just that of explaining why the statistical correlation (of smoking with possession of the gene) doesn't translate into an evidential dependency, from the agent's point of view.
- That's the puzzle the "Tickle Defence" and its descendants address.



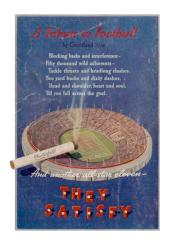
### Causal information is evidential information

### My response

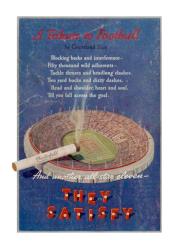
- When we're told the causal structure of the case (in an ordinary case like this!) we're told the relevant evidential dependencies – after all, that's what causation is!
- The puzzle is just that of explaining why the statistical correlation (of smoking with possession of the gene) doesn't translate into an evidential dependency, from the agent's point of view.
- That's the puzzle the "Tickle Defence" and its descendants address.



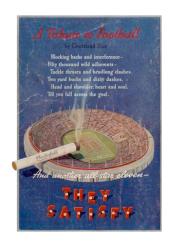
- Someone might come up with a (far less realistic) example, immune from anything like the Tickle Defence – i.e., a case in which there are genuine evidential dependencies, from the agent's point of view.
- If so, then I say that that's a case like the original Newcomb Problem, in which the causal structure isn't what Two Boxers take in the



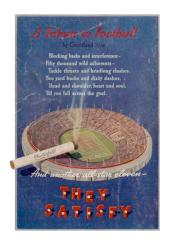
- Someone might come up with a (far less realistic) example, immune from anything like the Tickle Defence – i.e., a case in which there are genuine evidential dependencies, from the agent's point of view.
- If so, then I say that that's a case like the original Newcomb Problem, in which the causal structure isn't what Two Boxers take it to be



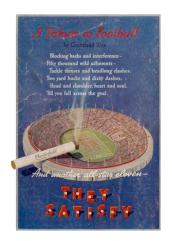
- Someone might come up with a (far less realistic) example, immune from anything like the Tickle Defence i.e., a case in which there *are* genuine evidential dependencies, from the agent's point of view.
- If so, then I say that that's a case like the original Newcomb Problem, in which the causal structure isn't what Two Boxers take it



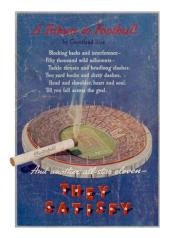
- Someone might come up with a (far less realistic) example, immune from anything like the Tickle Defence i.e., a case in which there *are* genuine evidential dependencies, from the agent's point of view.
- If so, then I say that that's a case like the original Newcomb Problem, in which the causal structure isn't what Two Boxers take it to be.



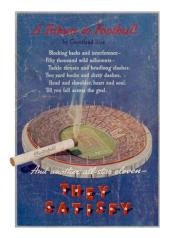
- Q: How do we know in advance whether we can just rely on our naïve causal intuitions?
- A: We don't, in general, and the fall-back is always "Try it and see!"
- This gives the right answer, from an EDTer's perspective, in both the cases already mentioned: you should one-box in the classic Newcomb problem, but two-box – i.e., smoke – in the medical problem.



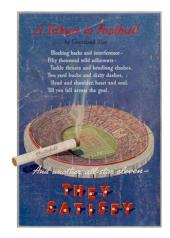
- Q: How do we know in advance whether we can just rely on our naïve causal intuitions?
- A: We don't, in general, and the fall-back is always "Try it and see!"
- This gives the right answer, from an EDTer's perspective, in both the cases already mentioned: you should one-box in the classic Newcomb problem, but two-box i.e., smoke in the medical problem.



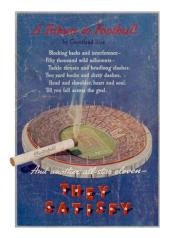
- Q: How do we know in advance whether we can just rely on our naïve causal intuitions?
- A: We don't, in general, and the fall-back is always "Try it and see!"
- This gives the right answer, from an EDTer' perspective, in both the cases already mentioned: you should one-box in the classic Newcomb problem, but two-box i.e., smoke in the medical problem.



- Q: How do we know in advance whether we can just rely on our naïve causal intuitions?
- A: We don't, in general, and the fall-back is always "Try it and see!"
- This gives the right answer, from an EDTer's perspective, in both the cases already mentioned: you should one-box in the classic Newcomb problem, but two-box —
  i.e., smoke in the medical problem.



- Q: How do we know in advance whether we can just rely on our naïve causal intuitions?
- A: We don't, in general, and the fall-back is always "Try it and see!"
- This gives the right answer, from an EDTer's perspective, in both the cases already mentioned: you should one-box in the classic Newcomb problem, but two-box i.e., smoke in the medical problem.



- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

True, one might prefer a more restrictive view of causation – e.g., a local notion, in the QM case – but in that case one should abandon both the First Principle and CDT.

- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the OM case – but in that case one should abandon both the First Principle and CDT.

- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the OM case – but in that case one should abandon both the First Principle and CDT.



- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the QM case – but in that case one should abandon both the First Principle and CDT.

- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the QM case – but in that case one should abandon both the First Principle and CDT.

- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the QM case – but in that case one should abandon both the First Principle and CDT.

- A large part of the apparent "hardness" of Newcomb problems stems from unquestioned acceptance, on both sides, of a view about the "true" causal structure of the case.
- Once we ditch that view, recognising the link between causal structure and a pre-causal notion of effective strategy, the fog clears.<sup>3</sup>
- That option aside, CDT and EDT now the same thing! now agree that we should one-box in the classic case.
- Lingering doubt about that is lingering attachment to false metaphysics.
- EDT is not only compatible with CDT, but provides the only viable form of CDT.

<sup>&</sup>lt;sup>3</sup>True, one might prefer a more restrictive view of causation – e.g., a *local* notion, in the QM case – but in that case one should abandon both the First Principle and CDT.

## One World One Dream One Box

## One World One Dream One Box

