The Logic of Counterfactuals and the Epistemology of Causal Inference

A Dose of Econometrics for Everyone

Hanti Lin
ika@ucdavis.edu

UC DAVIS
UNIVERSITY OF CALIFORNIA

Presented at EpiCause 2024 on May 3

The Topic for Today

- There is is very influential framework for causal inference in health and social sciences: the **Rubin causal model** (Rubin 1974).
- But that framework raises a worry: it assumes a controversial logical principle called **Conditional Excluded Middle** (Dawid 2000).
- In reply, I will argue that the Rubin causal model can receive an update that dispenses with that logical principle.
- This will be done while preserving an important fruit of the framework: instrumental variable estimation of **LATE** (Imbens & Angrist 1994).

1. Introducing the Controversy

Counterfactual Variables in the Rubin Causal Model

$$Cured_i^{t=1} = \ldots$$

Counterfactual Variables in the Rubin Causal Model

$$Cured_i^{t=1} = 1$$

Counterfactual Variables in the Rubin Causal Model

$Cured_i^{t=1} = 1$ means that,
if individual i took the treatment ($Take$ =1),
i would be cured ($Cured$ = 1).

Counterfactual Variables in the Rubin Causal Model

$Cured_i^{t=1} = 1$ means that,
if individual i took the treatment ($Take = 1$),
i would be cured ($Cured = 1$).

$Cured_i^{t=0} = 1$ means that,
if individual i did not take the treatment ($Take = 0$),
i would be cured ($Cured = 1$).

An Assumption: Conditional Excluded Middle (CEM)

Consider an individual i who actually does not take the treatment.
Then *Conditional Excluded Middle* says the following:

CEM	$Cured_i^{t=1}$	=	either 1 or 0.
	The result that i would have if i took the treatment	is	either being cured, or not being cured.

Either [if i took the treatment, i would be cured],
or [if i took the treatment, i would not be cured].

Dawid's (2000) Argument Against CEM

CEM	$Cured_i^{t=1}$	=	either 1 or 0.
	The result that *i* would have if *i* took the treatment	is	either being cured, or not being cured.

Either [if *i* took the treatment, *i* would be cured], or [if *i* took the treatment, *i* would not be cured].

The statistician Dawid (2000) raises a worry:

- In an indeterministic world, there is no such thing as **the** result that *i* would have if *i* took the treatment. For *i* could be cured, and could be not cured.
- So, assuming CEM is assuming a kind of fatalism/determinism.
- Even some proponents of the Rubin causal model agree that this is a problem, such as Robins and Greenland (2000).

CEM has been quite controversial in philosophy of language since 1970's.

Lewis' (1973) Argument Against CEM

CEM Either (*A*) if *i* took the treatment, *i* would be cured, or (*B*) if *i* took the treatment, *i* would not be cured.

(1) Suppose that we are in an indeterministic world.

(2) So, if *i* took the treatment, *i* would have a nonzero probability to be cured, and would have a nonzero probability to be not cured.

(3) So, if *i* took the treatment, *i* could be cured, and could be not cured.

> (4) Suppose for *reductio* that (*B*) holds,
> namely, if *i* took the treatment, *i* would not be cured.
>
> (5) Then, by (3) and (4), we have:
> if *i* took the treatment, *i* would not be cured and could be cured,
> which is absurd.

(6) So, by the *reductio* argument (4)-(5), it follows that (*B*) is false.

(7) By the same argument, (*A*) is false, too.

A Potentially Powerful Argument for CEM

A very large group of philosophers—those in the naturalist tradition—generally take seriously a type of argument.

The idea is that the success of our best scientific theory *T* provides good reason for us to believe in the assumptions that are indispensable in that theory *T*.

A Potentially Powerful Argument for CEM

CEM is assumed, and seems to be **indispensable**, in our best theory of causal inference in health and social sciences—the theory that led to one half of the 2021 Nobel Prize in Economics (Imbens and Angrist 1994).

So, it seems that we should accept CEM.

Plan for Today

I love the Rubin causal model and its applications. But I am skeptical of CEM.

I will give the Rubin causal model an update,
a **fully stochastic update** that

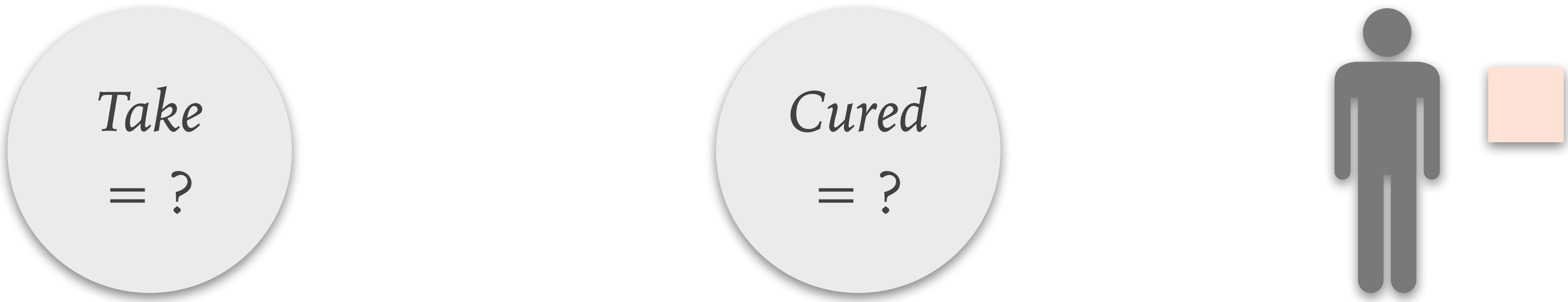
- preserves the Nobel prize winning application (i.e. estimation of LATE)
- dispenses with CEM.

To that end, many key ideas from the Rubin causal model will be integrated into

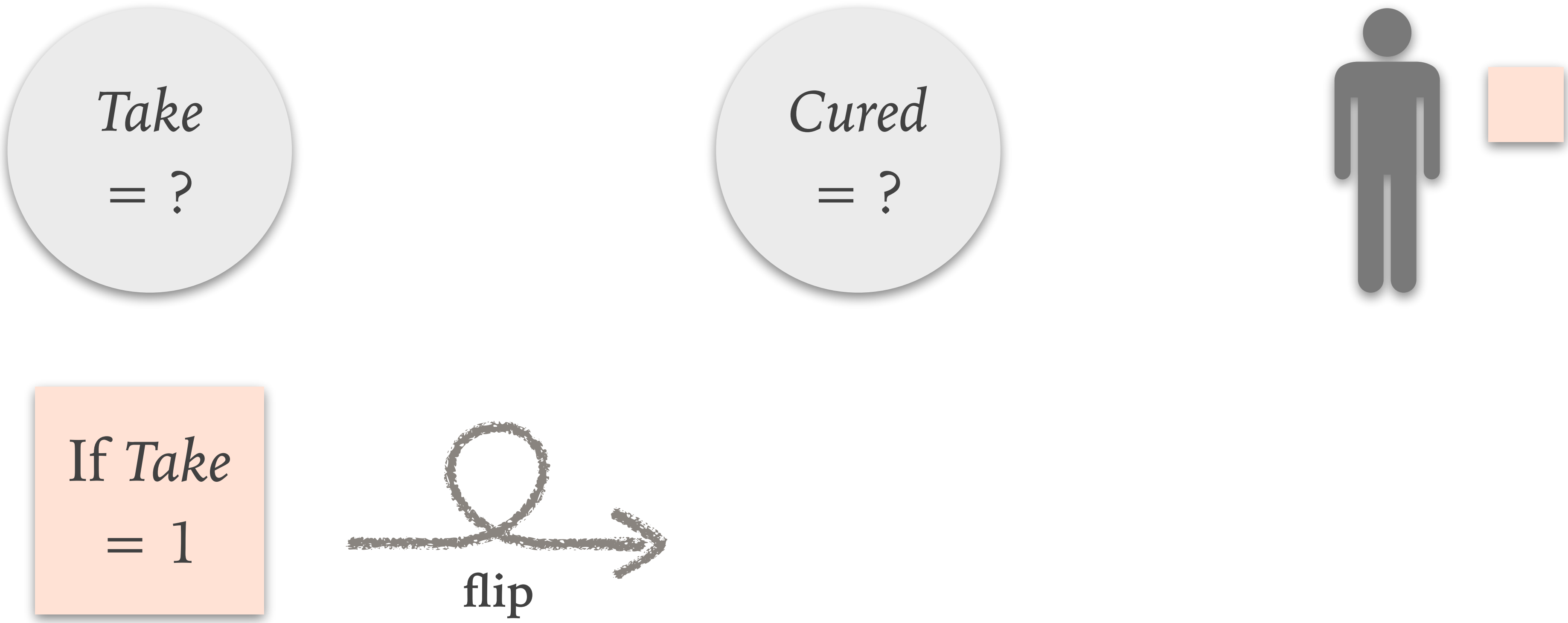
- a causal Bayes net,
- *not* to be confused with Pearl's structural equation model, which *still* assumes CEM.

2. The Rubin Causal Model:
A Crash Course

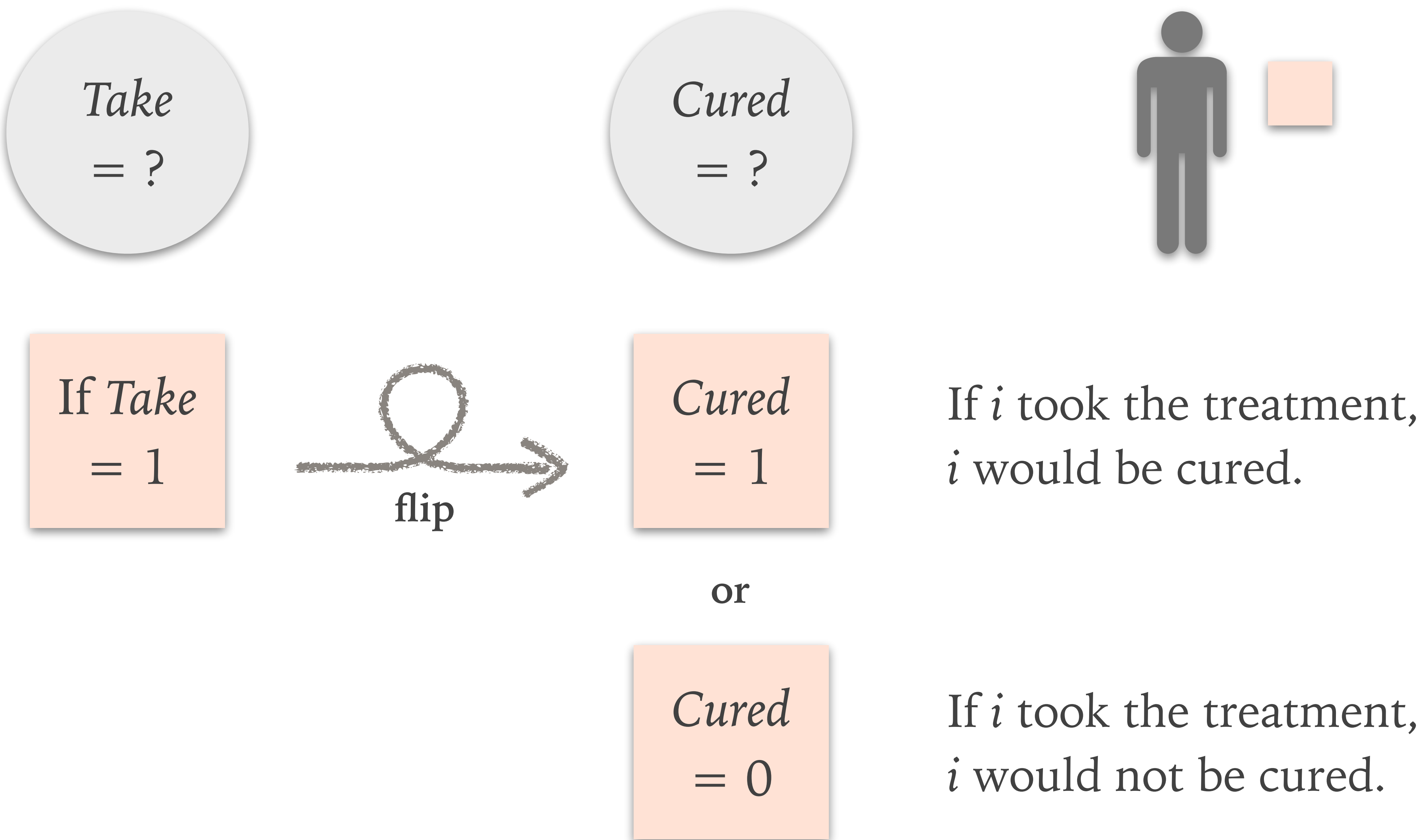
What If One Took the Treatment?



What If One Took the Treatment?



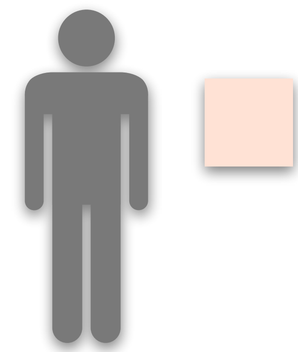
What If One Took the Treatment?



What If One Took the Treatment?

Take = ?

Cured = ?



If *Take* = 1 —flip→ *Cured* = 1

If i took the treatment, i would be cured.

or

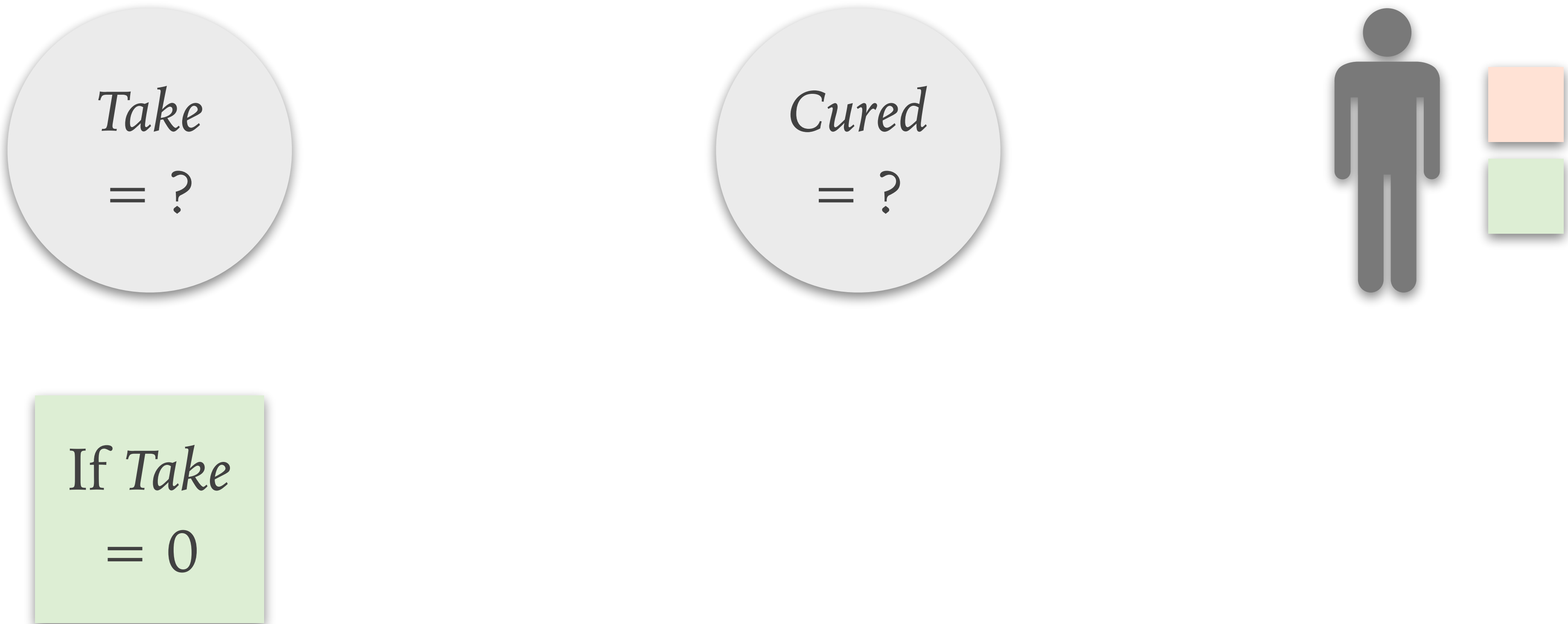
Cured = 0

If i took the treatment, i would not be cured.

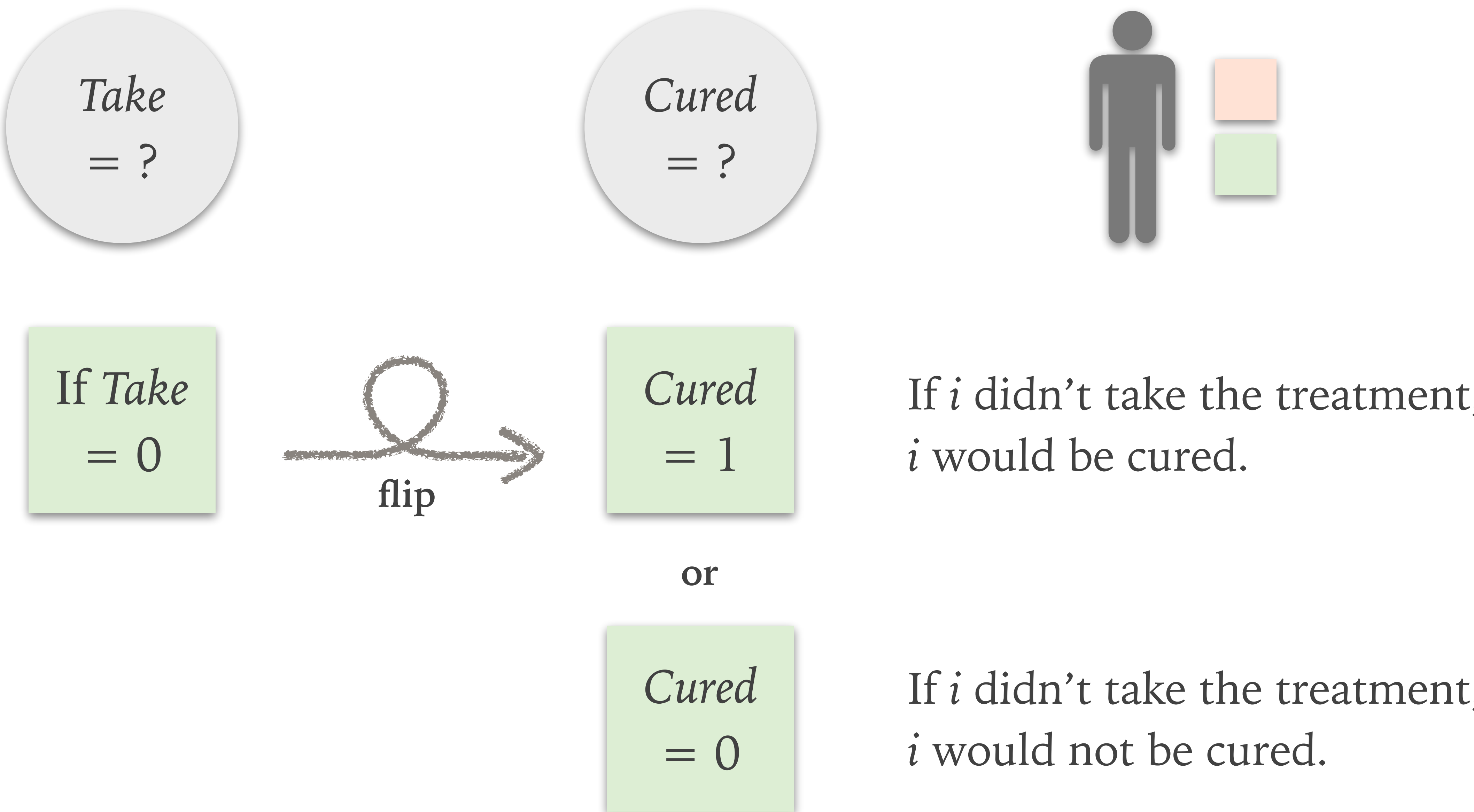
CEM is built in.

In symbol, $Cured_i^{t=1} = 1$ or 0.

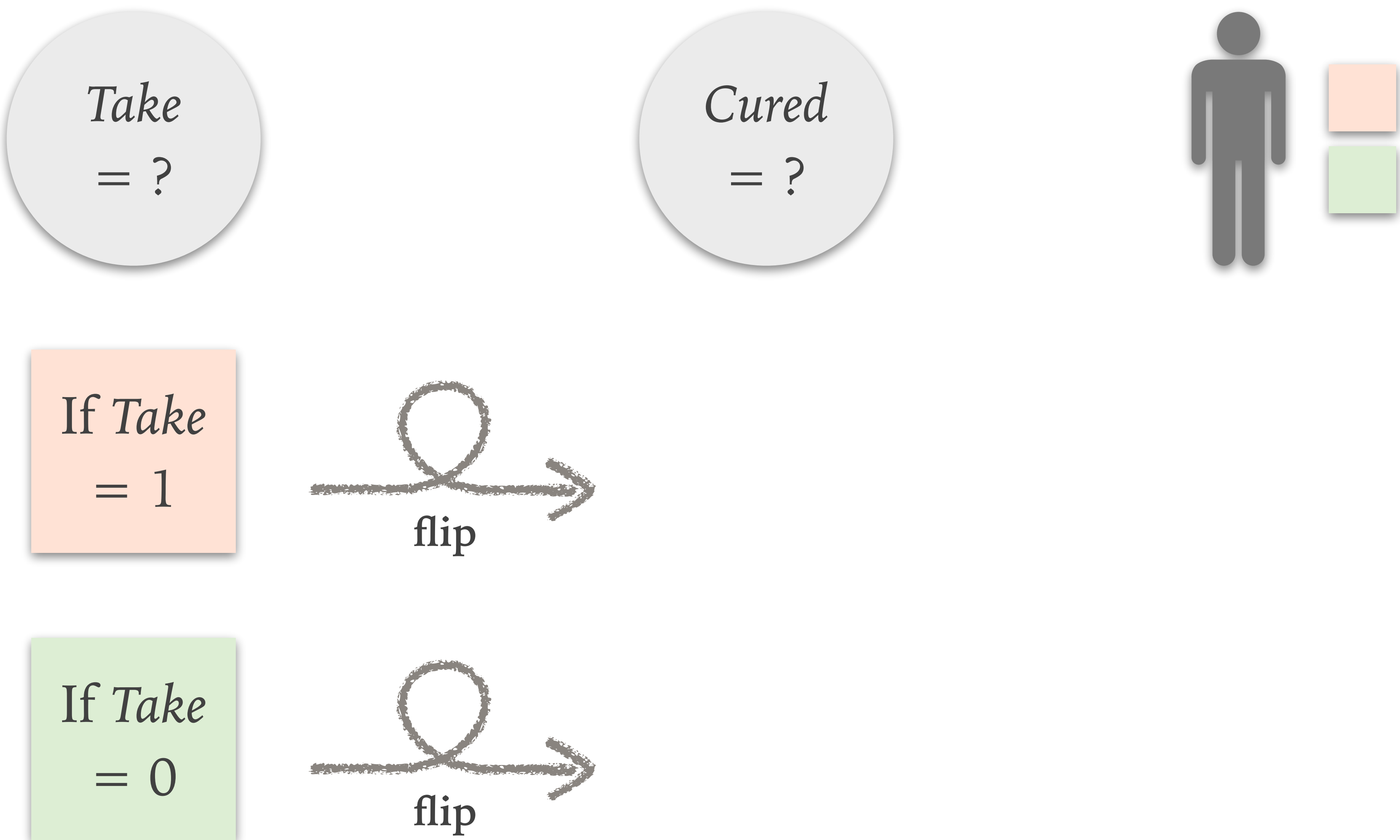
What If One *Didn't* Take the Treatment?



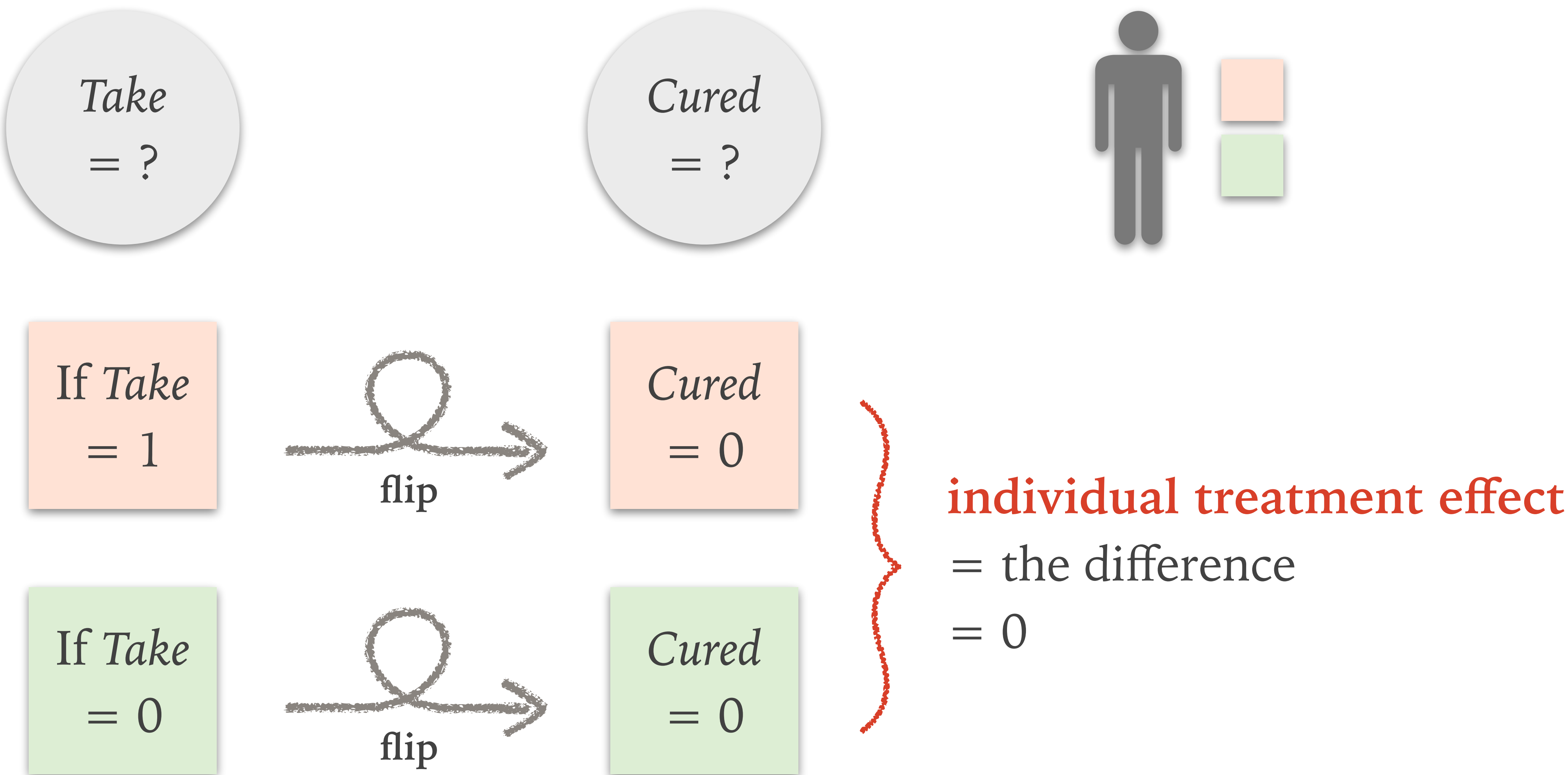
What If One *Didn't* Take the Treatment?



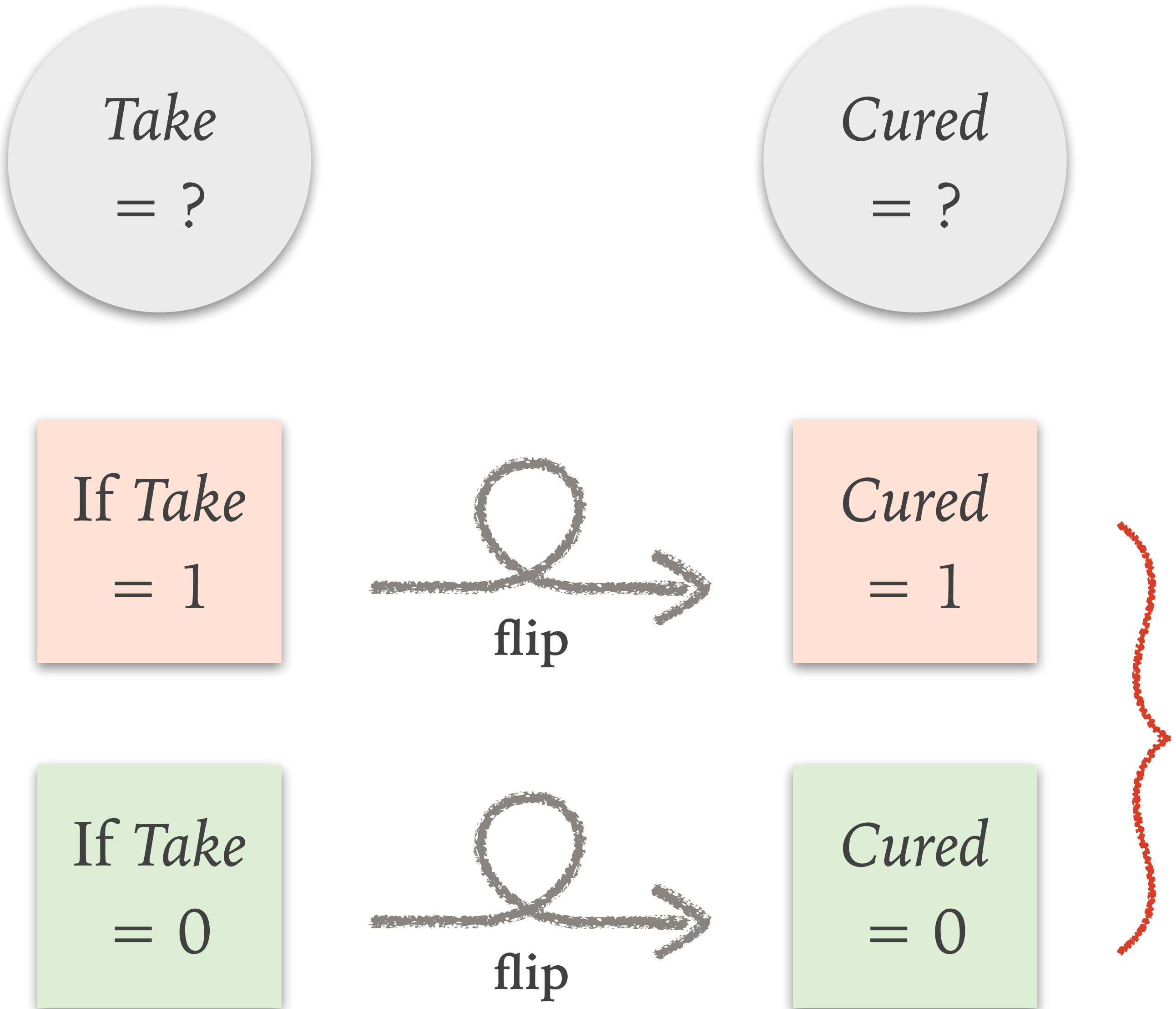
Def: Individual Treatment Effect (ITE)

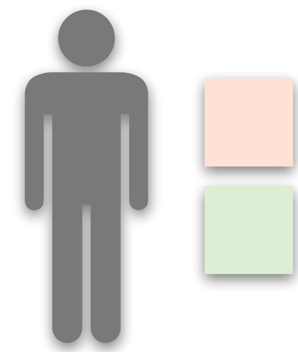


Def: Individual Treatment Effect (ITE)



Def: Individual Treatment Effect (ITE)



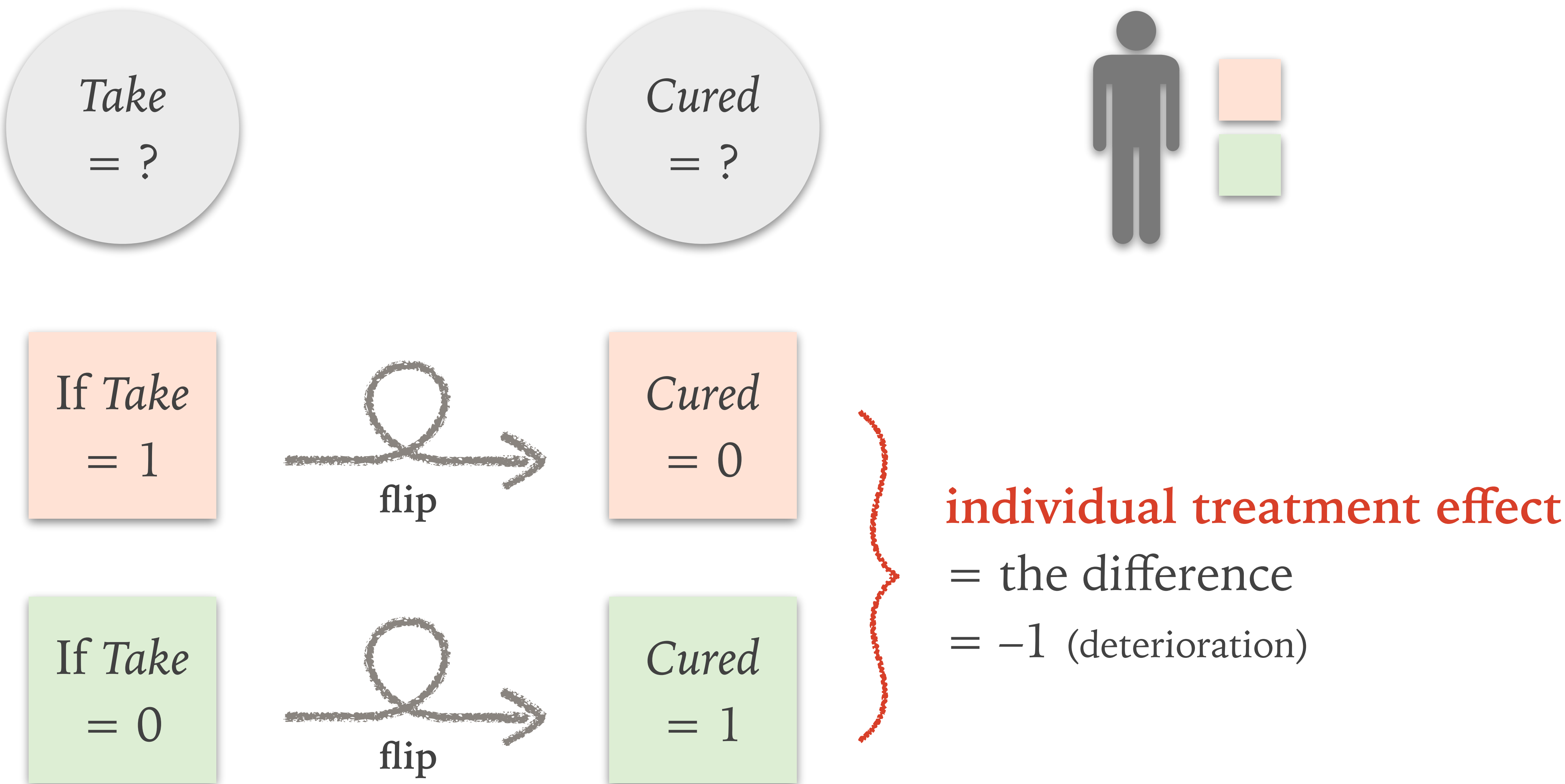


individual treatment effect

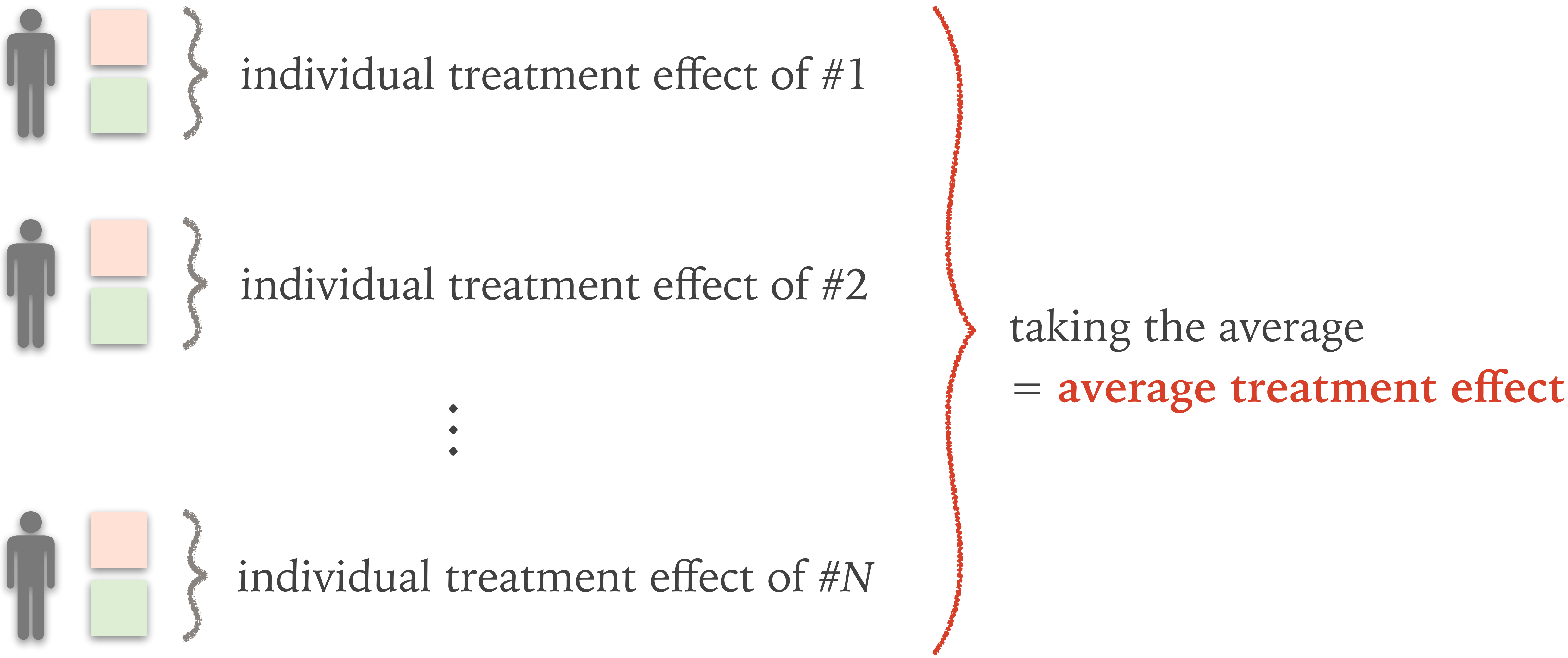
= the difference

= 1 (improvement)

Def: Individual Treatment Effect (ITE)



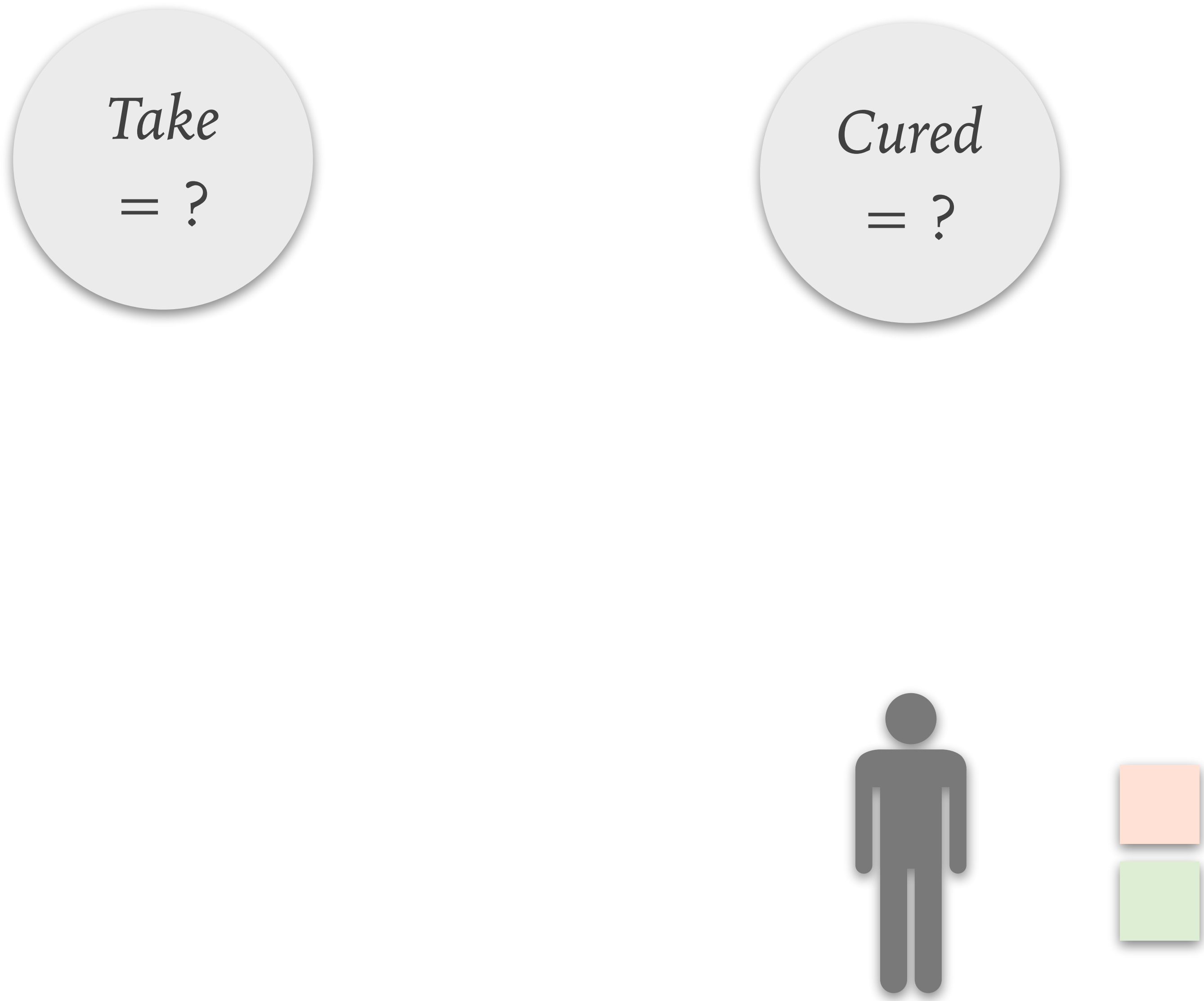
Def: Average Treatment Effect (ATE)



ATE can be easily estimated if we can perform a perfect RCT: if we can randomly select people and force each to flip a card according to the result of a coin toss. **Crux**: We often cannot do that!

3. LATE Comes to Rescue

Original Setup



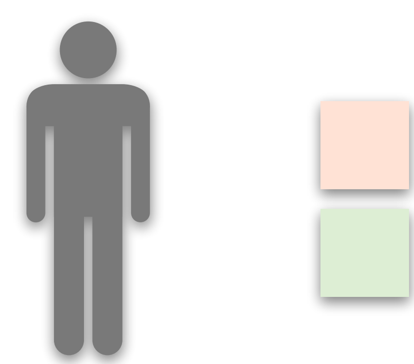
Add Assignment to Treatment/Control Group

Assign = ?

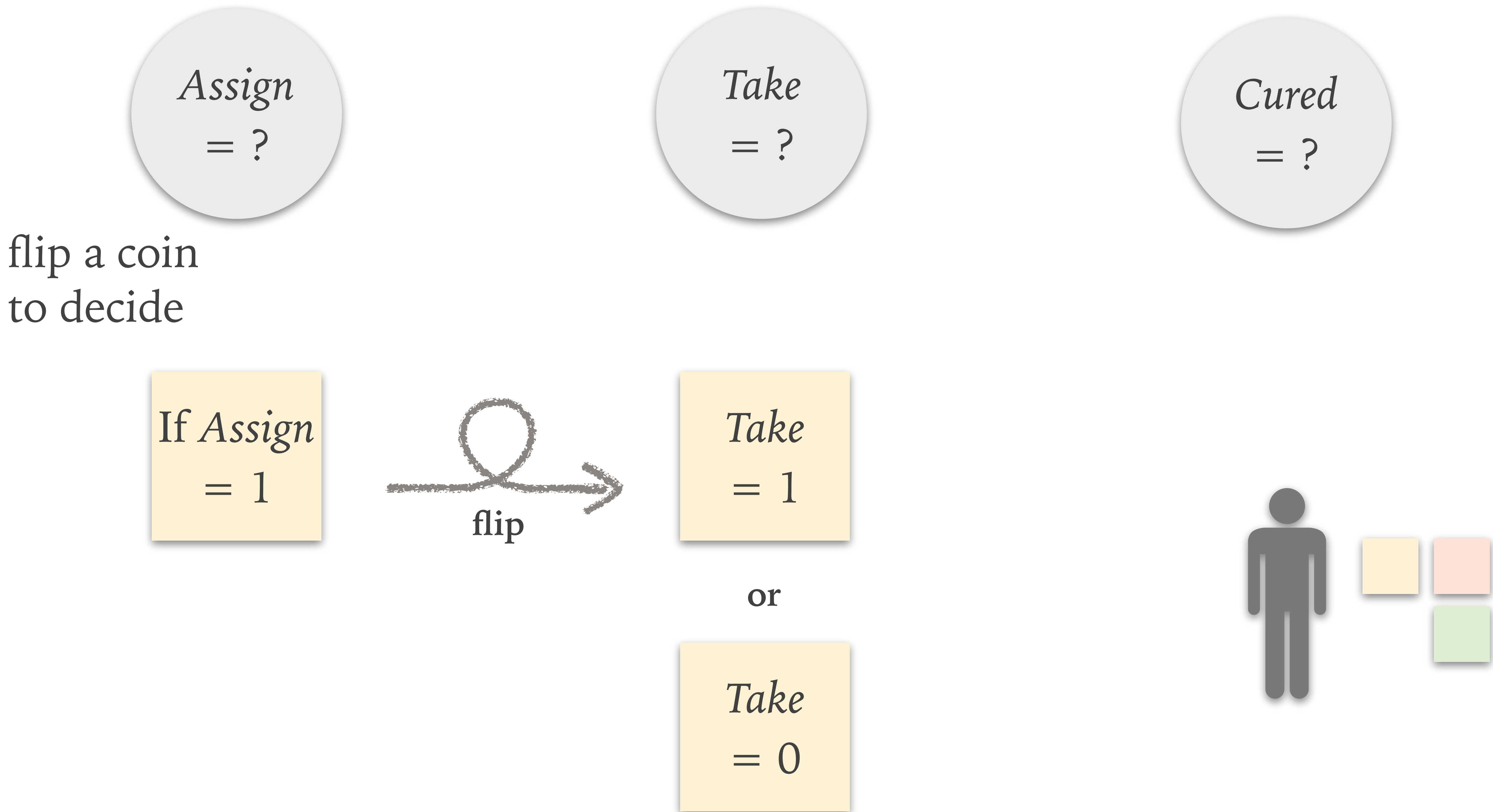
flip a coin
to decide

Take = ?

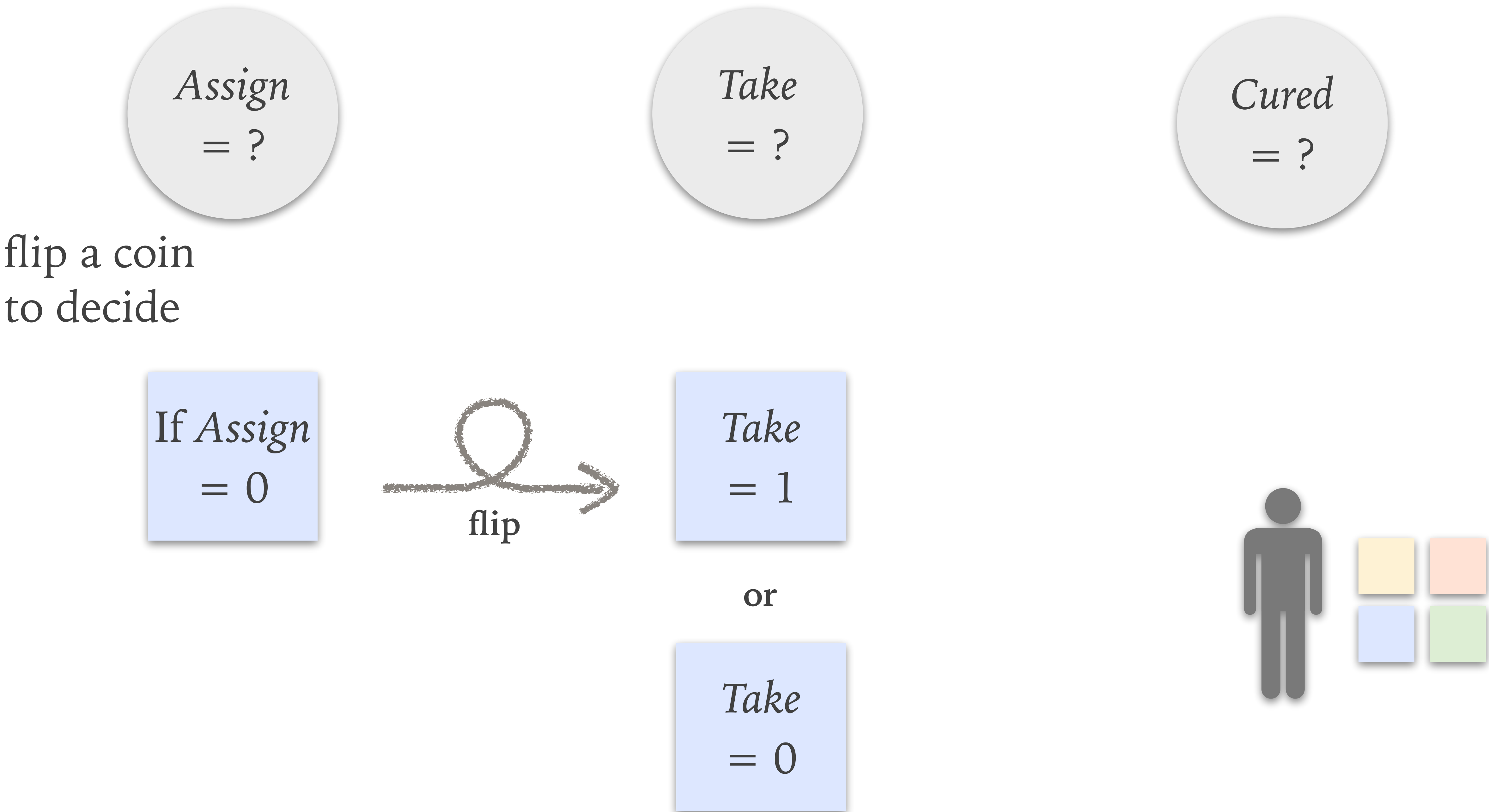
Cured = ?



Add a New Card:
What If One Were Assigned to the Treatment Group



Add a New Card: What If One Were Assigned to the Control Group



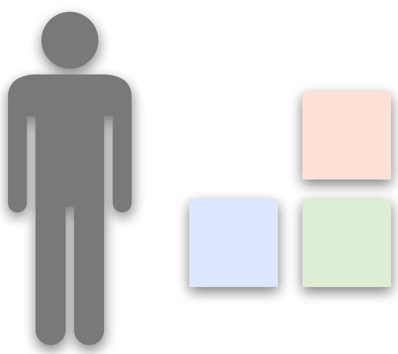
How the Game Plays Out

How the Game Plays Out

Assign = 1

Take = ?

Cured = ?



flip a coin
to decide

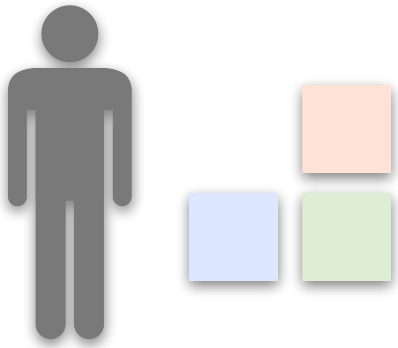
If *Assign* = 1

How the Game Plays Out

Assign = 1

Take = 0

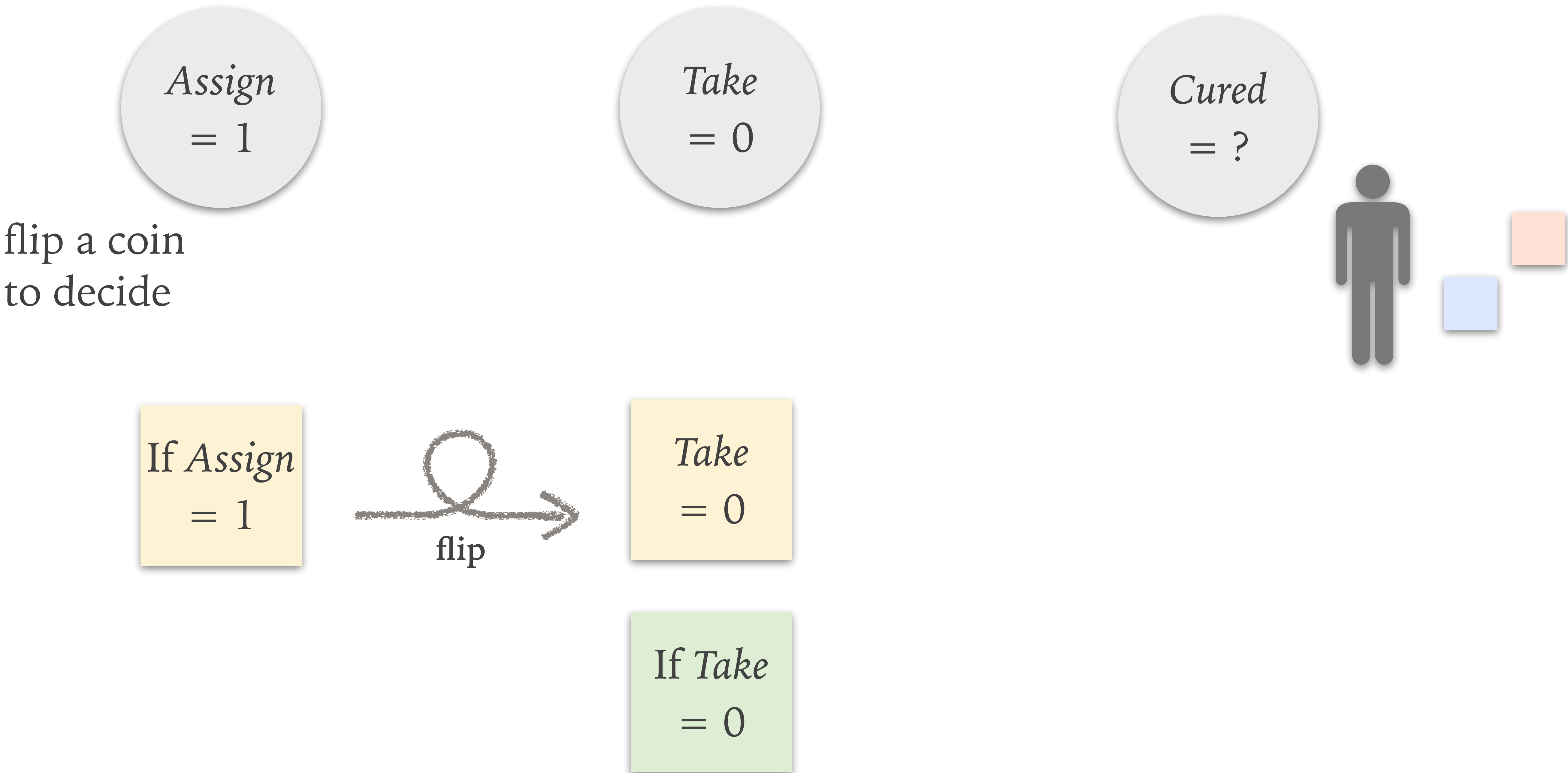
Cured = ?



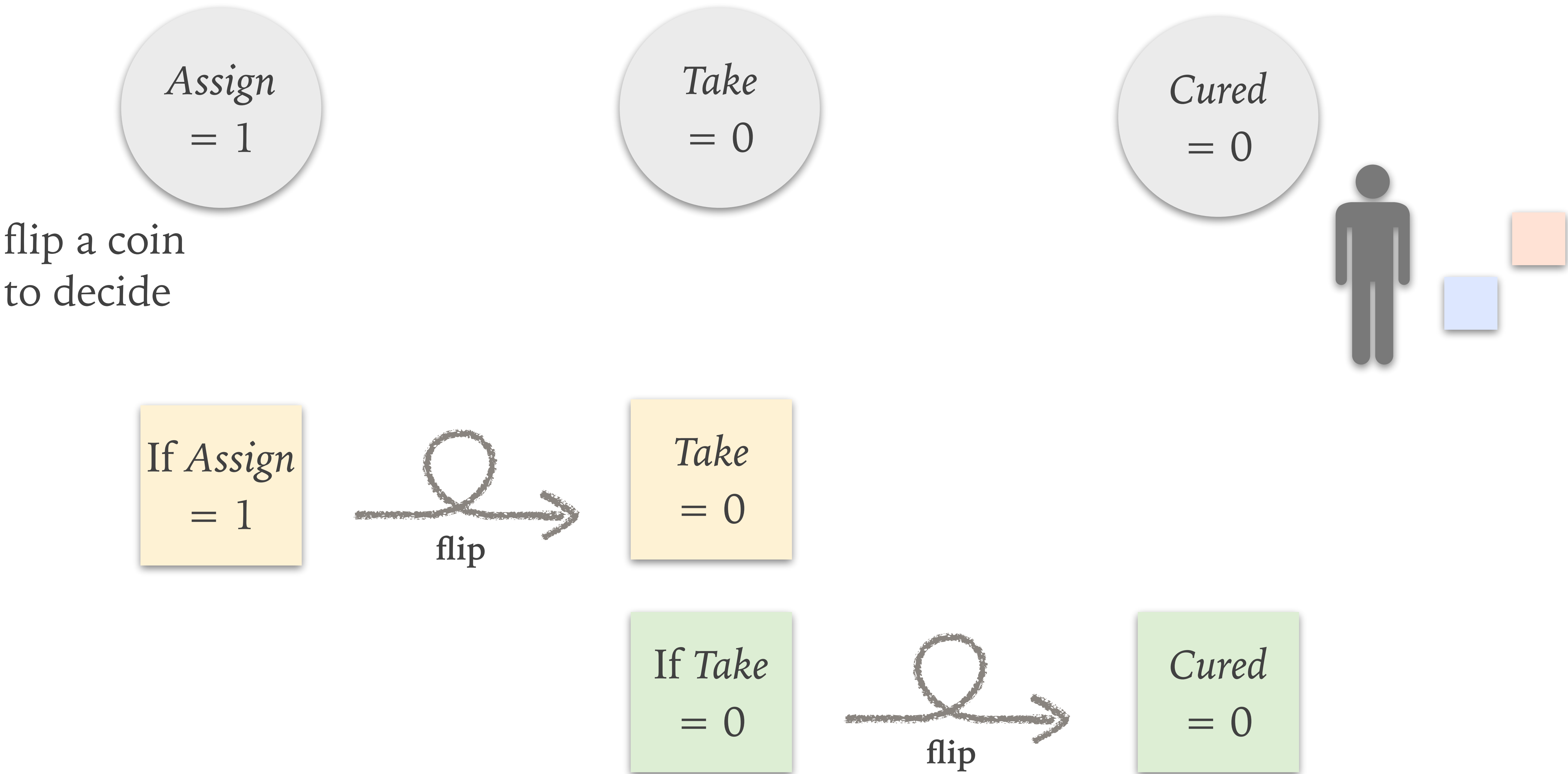
flip a coin
to decide

If *Assign* = 1 → flip → *Take* = 0

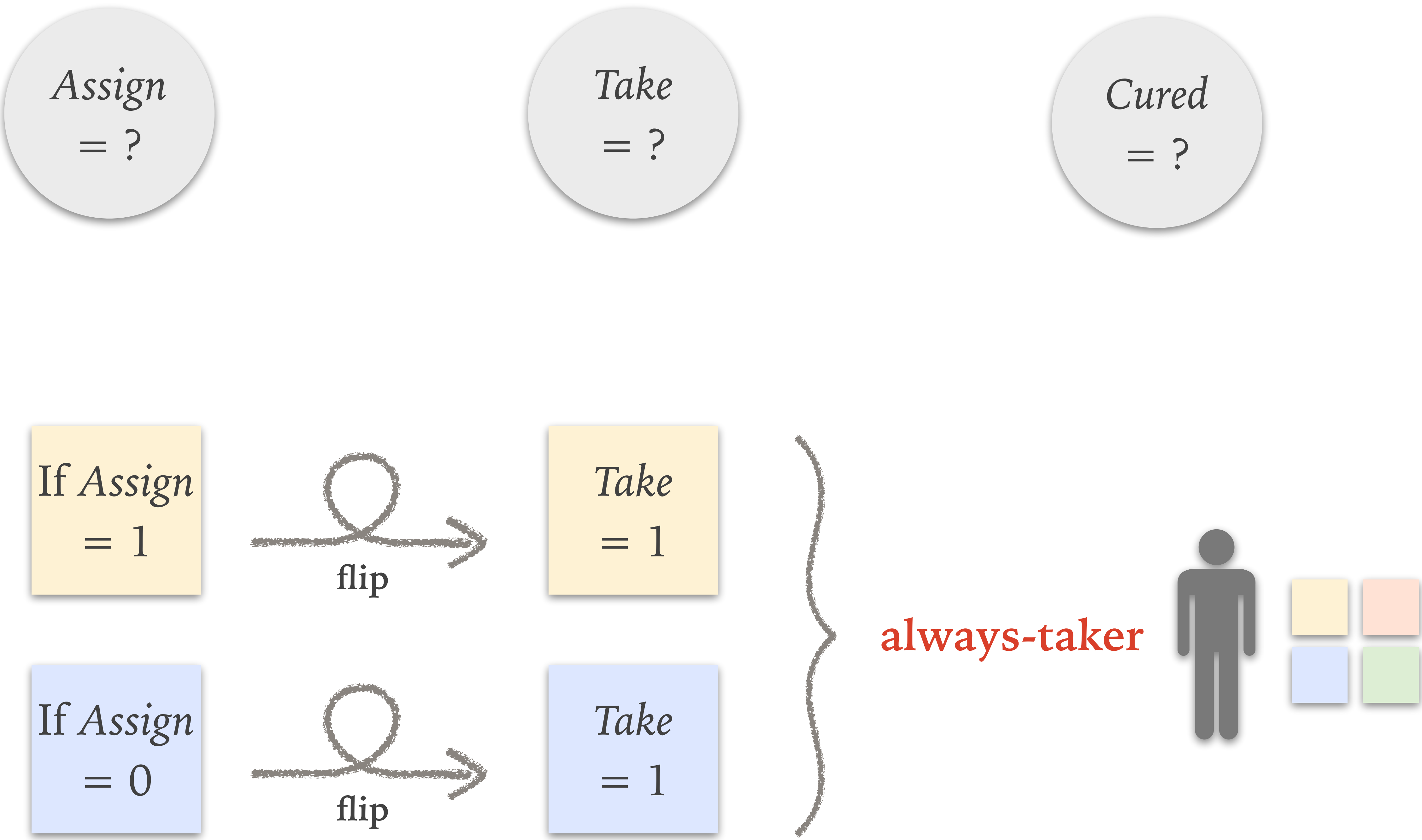
How the Game Plays Out



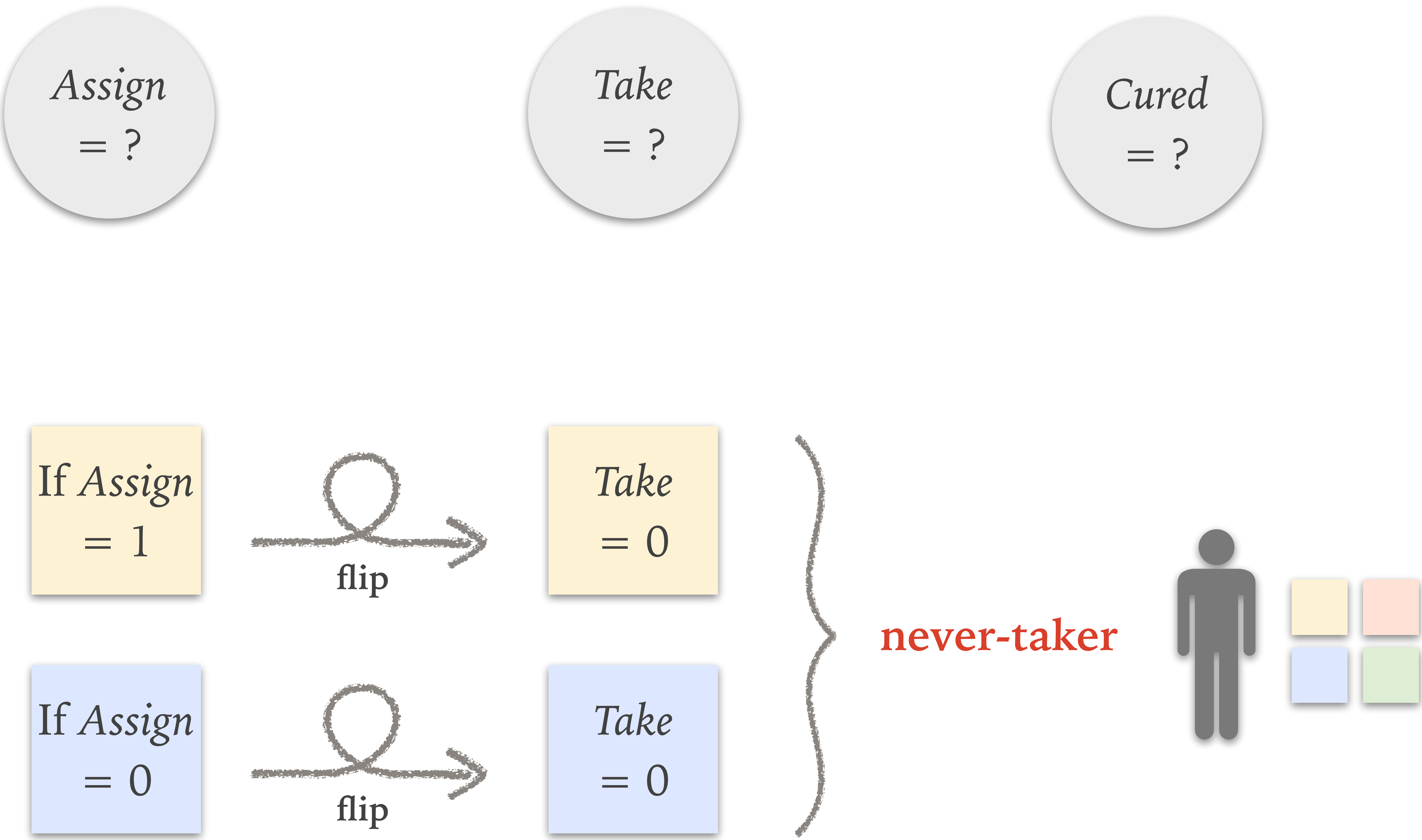
How the Game Plays Out



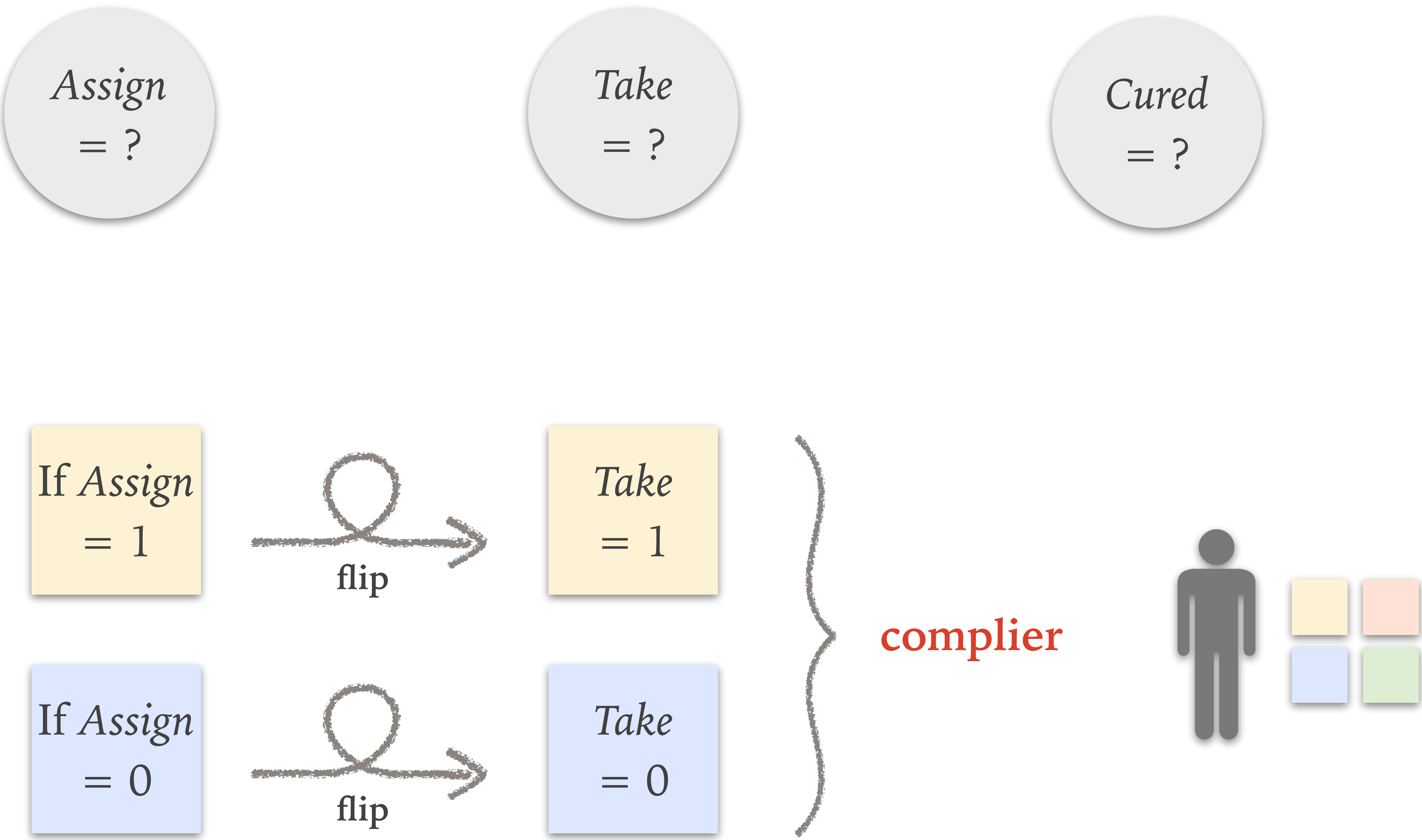
Def: Subpopulation 1, *Always-Takers*



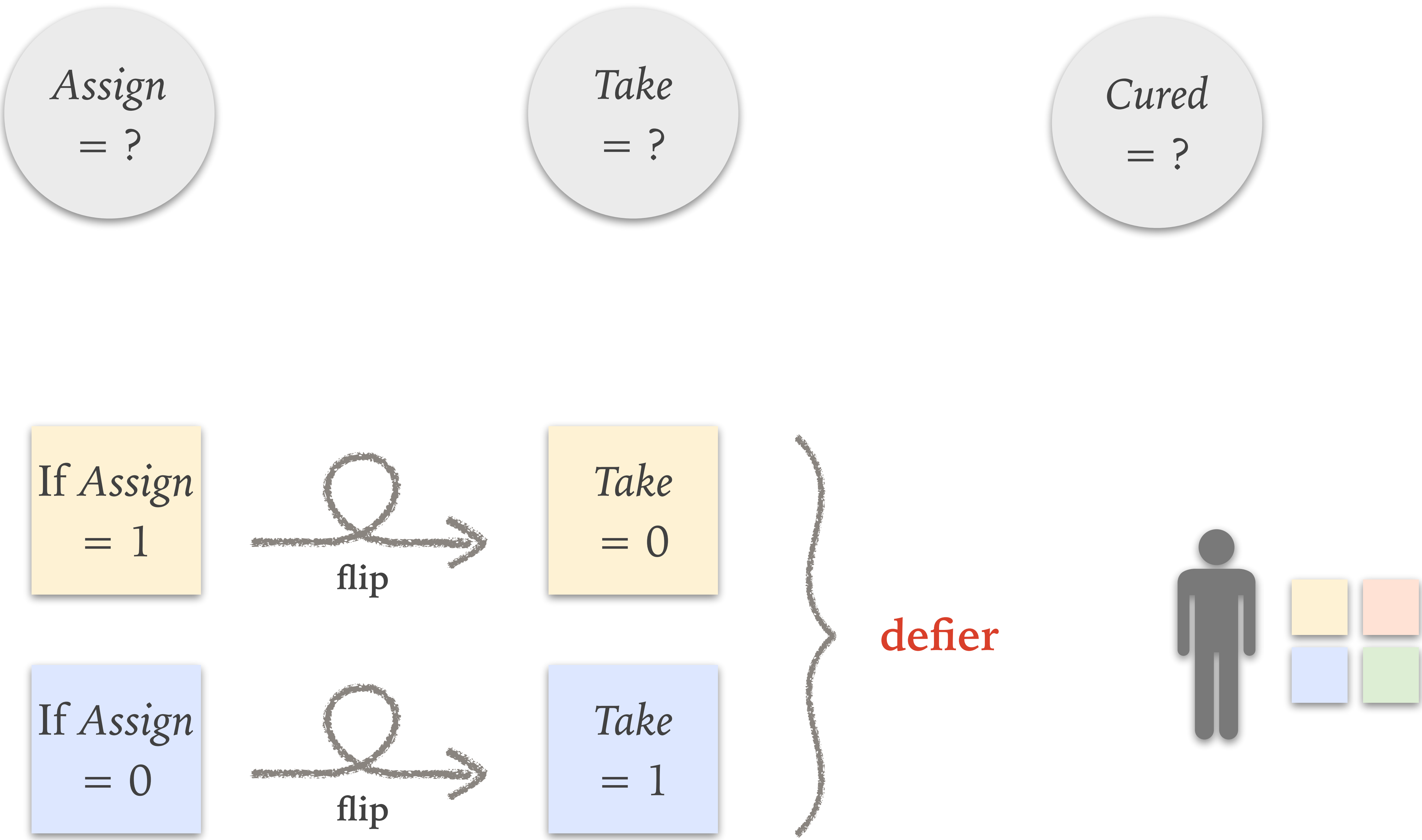
Def: Subpopulation 2, *Never-Takers*



Def: Subpopulation 3, *Compliers*



Def: Subpopulation 4, *Defiers*



Def: LATE

LATE, i.e.

Local
Average
Treatment
Effect
(of the subpopulation of compliers)

= the average of the individual treatment effects of the compliers

Identification Result (Imbens and Angrist 1994)

In this card game (in more general settings captured by the assumptions stated by Imbens and Angrist),

if people are randomly selected from the population and then assigned to the treatment/control group by flipping a coin,

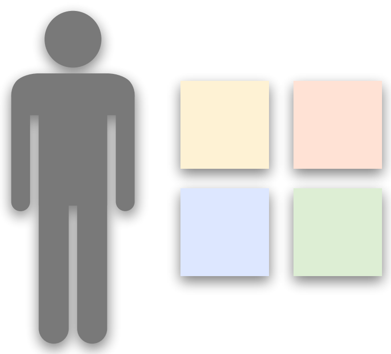
if there are no defiers,

then LATE can be expressed *solely* in terms some quantities that can be estimated *without* forcing anyone to take the treatment:

$$\text{LATE} = \frac{\Pr(Cured = 1 \mid Assign = 1) - \Pr(Cured = 1 \mid Assign = 0)}{\Pr(Take = 1 \mid Assign = 1) - \Pr(Take = 1 \mid Assign = 0)}$$

4. A Fully Stochastic Update

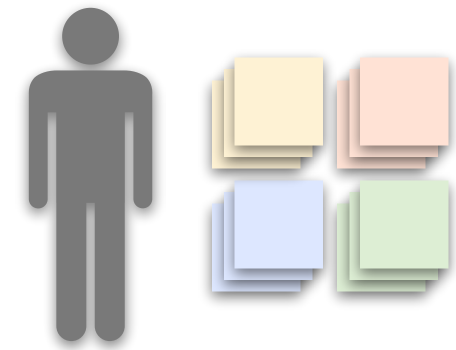
Original Setup

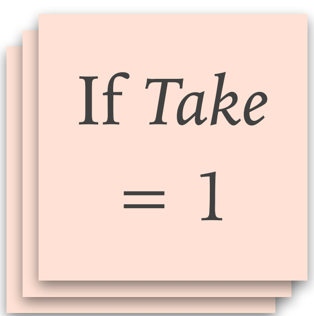


Single Cards → Decks of Cards

Use Decks to Get Rid of Determinism

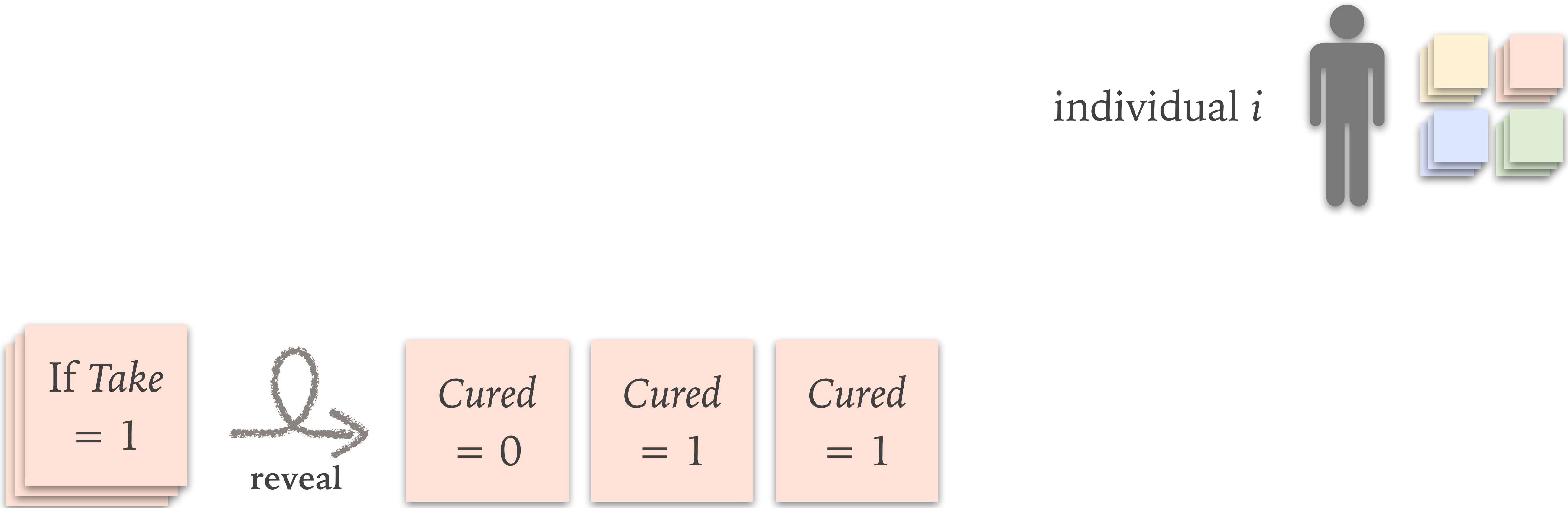
individual i





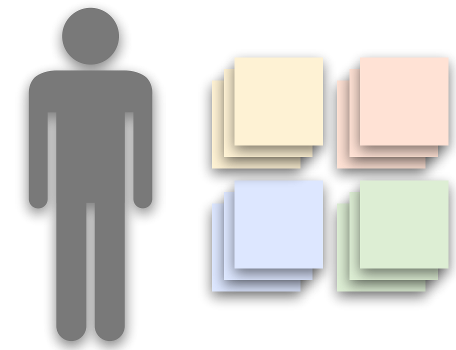
If i took the treatment, i would randomly draw a card from this deck to determine the medical result.

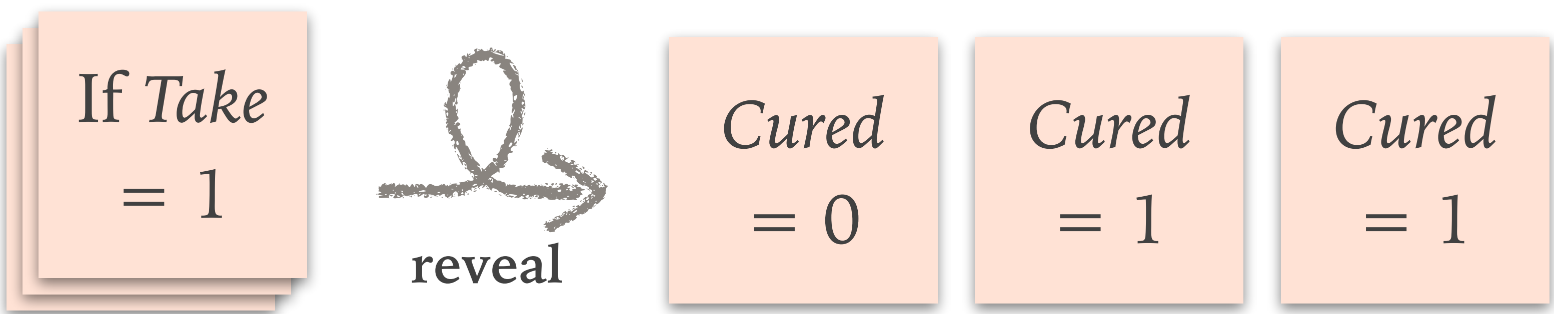
Use Decks to Get Rid of Determinism



Use Decks to Get Rid of Determinism

individual i





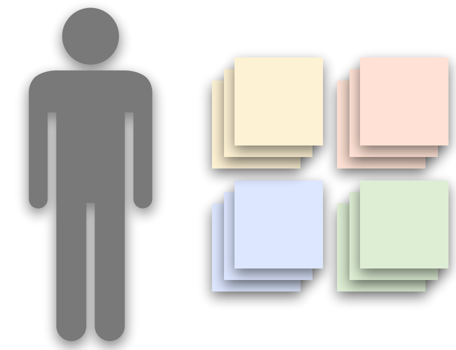
If i took the treatment, i would have a certain probability of being cured: 2/3.

The proportion of the "*Cured* = 1" cards in the deck for "If *Take* = 1" is equal to 2/3.

In symbol,
$\Pr_i^{t=1}(Cured = 1) \ = \ 2/3.$

Individual Treatment Effect (ITE) *Redefined*

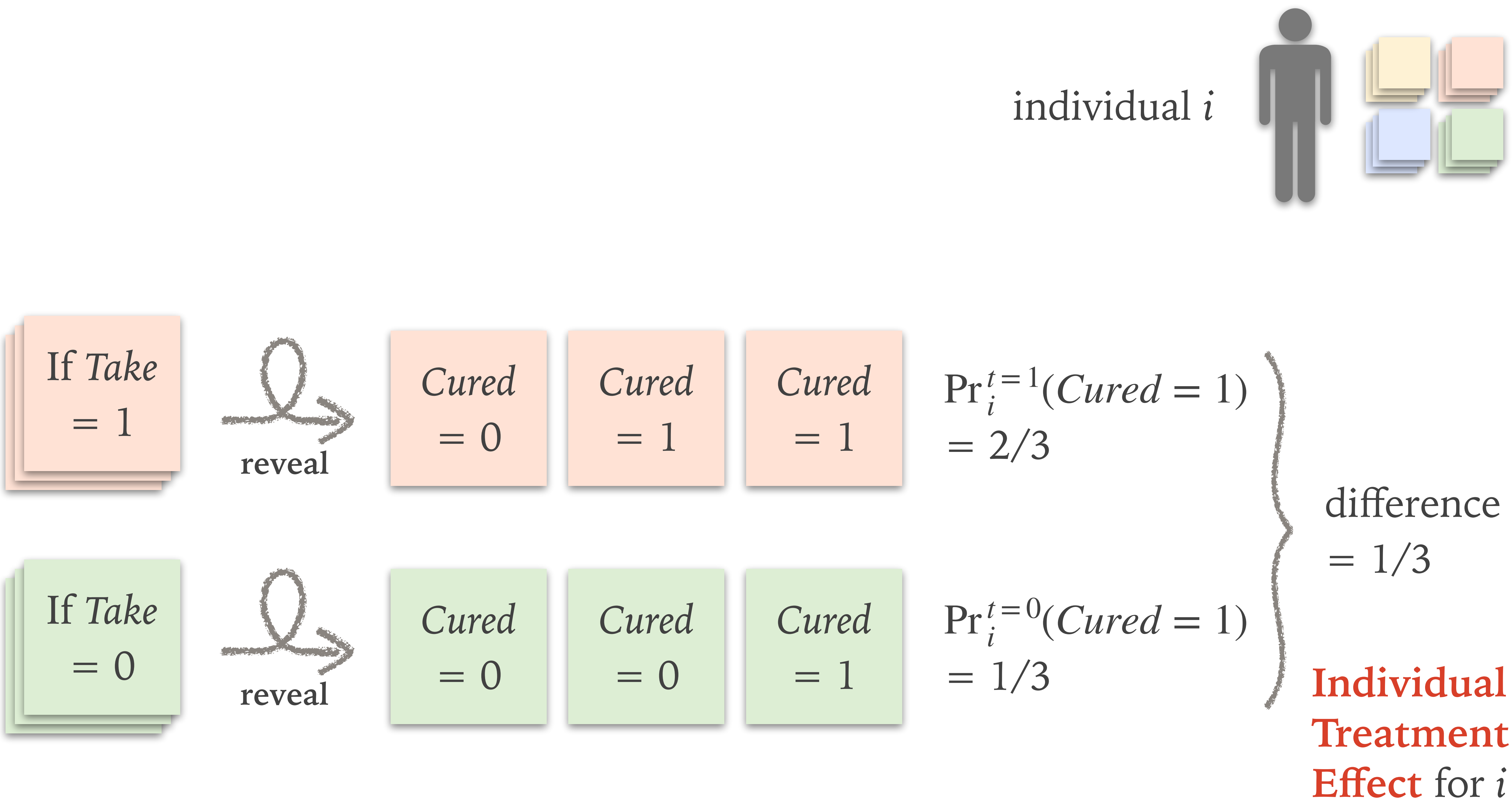
individual i



If *Take* = 1 → reveal → *Cured* = 0 | *Cured* = 1 | *Cured* = 1

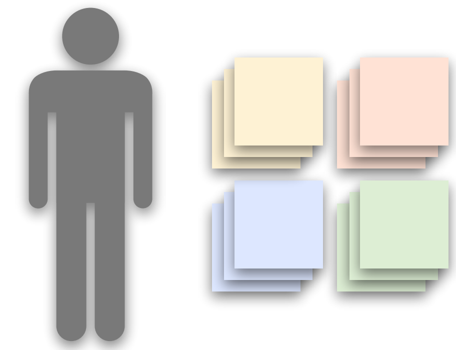
If *Take* = 0 → reveal → *Cured* = 0 | *Cured* = 0 | *Cured* = 1

Individual Treatment Effect (ITE) *Redefined*



Def: Degree of Compliance

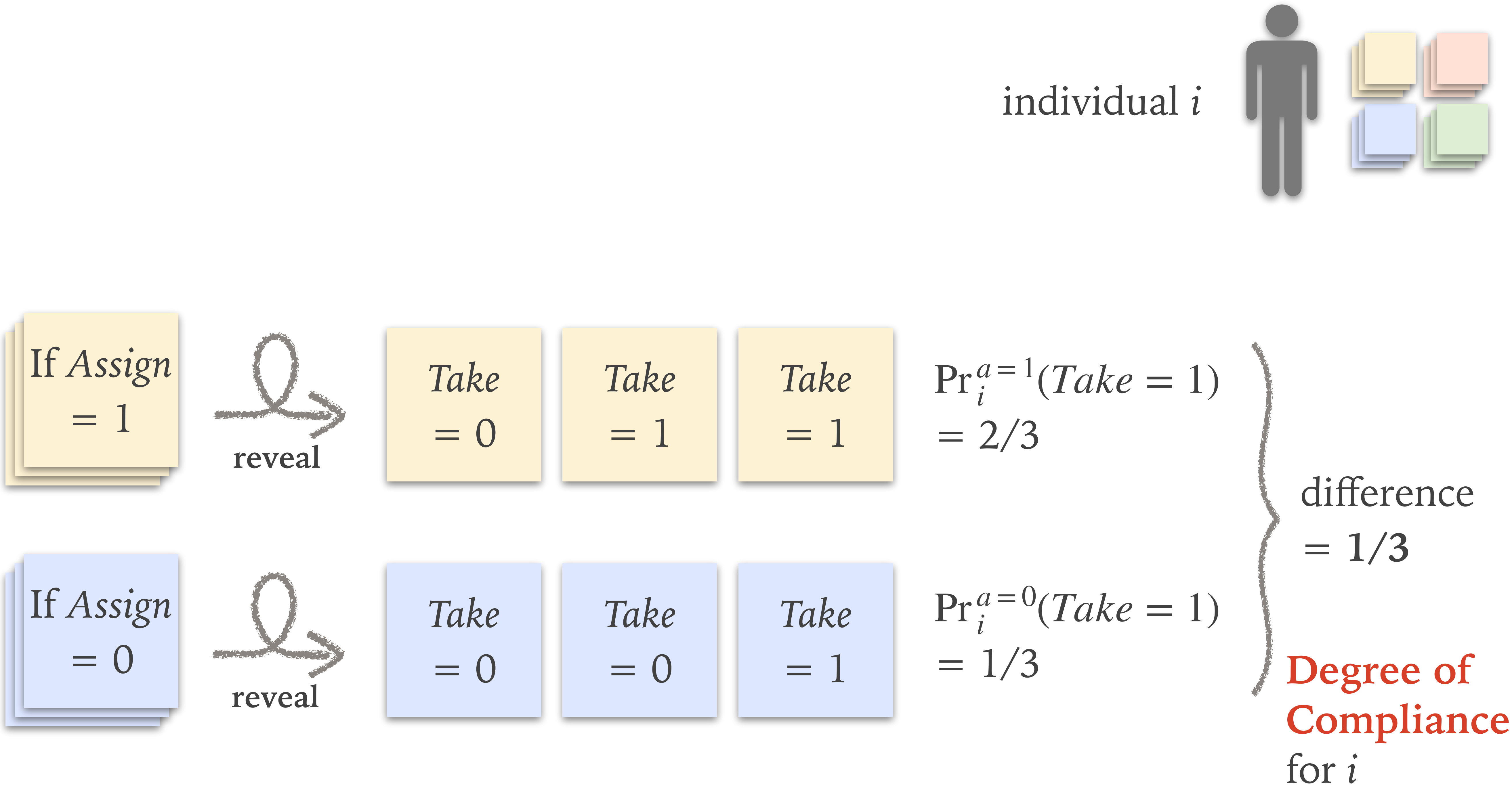
individual i



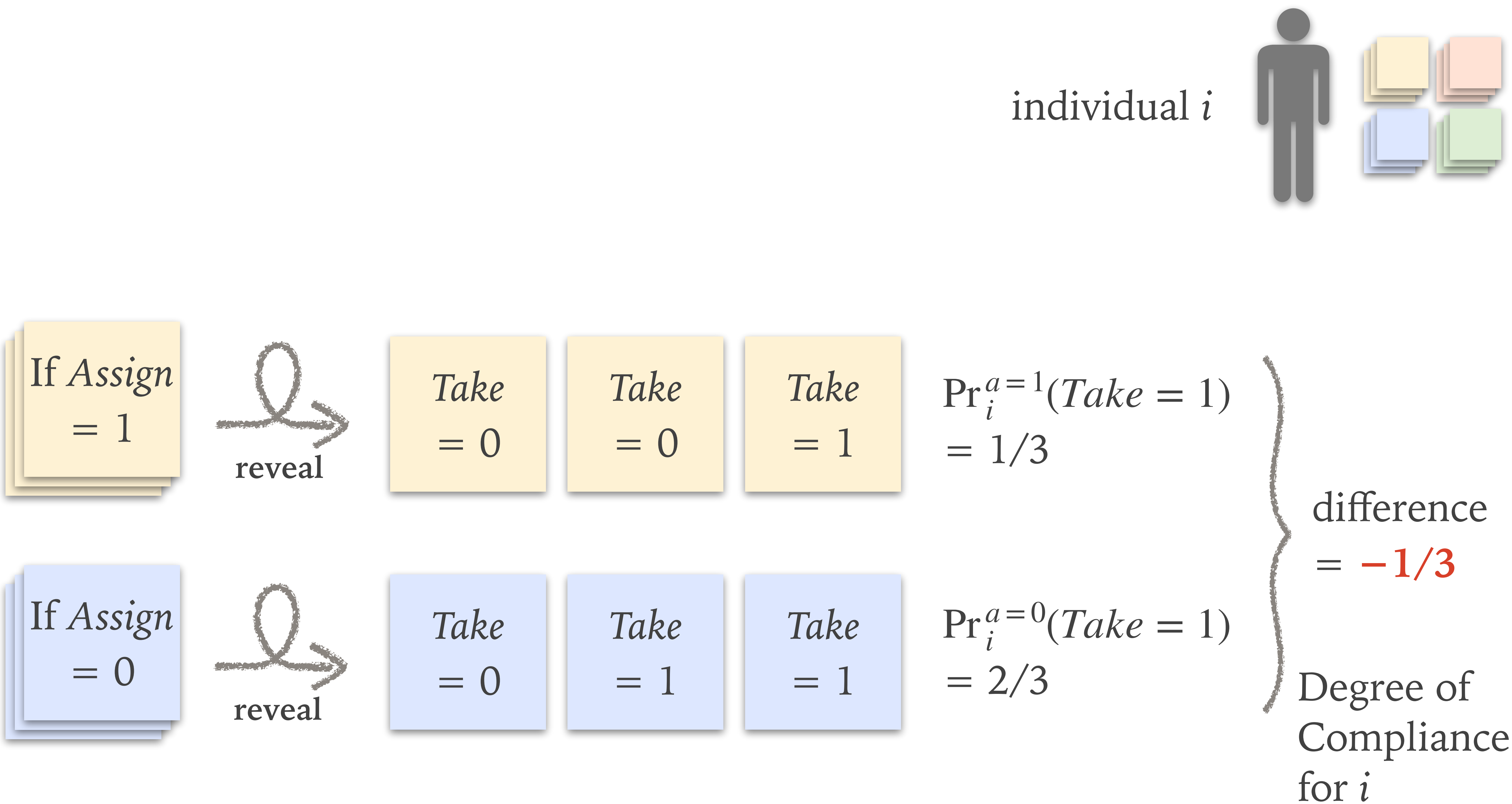
If *Assign* = 1 → reveal → *Take* = 0 | *Take* = 1 | *Take* = 1

If *Assign* = 0 → reveal → *Take* = 0 | *Take* = 0 | *Take* = 1

Def: Degree of Compliance



Degree of Compliance Can Be *Negative*



Defiers and Compliers *Refined*

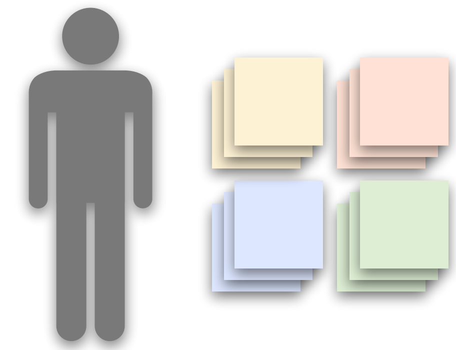
A **defier** is

- an individual with degree of compliance < 0.

A **complier** is

- an individual with degree of compliance > 0.

individual i



Def: DATE (A Generalization of LATE)

DATE, i.e.

Degree-of-compliance-weighted
Average
Treatment
Effect
of the subpopulation of compliers

= a weighted average of the individual treatment effects of the compliers, with the weights set to be degrees of compliance

Def: DATE (A Generalization of LATE)

DATE, i.e.

Degree-of-compliance-weighted
Average
Treatment
Effect
of the subpopulation of compliers

In the special case in which every deck turns out to be a single card, …

= a weighted average of the individual treatment effects of the compliers, with the weights set to be degrees of compliance

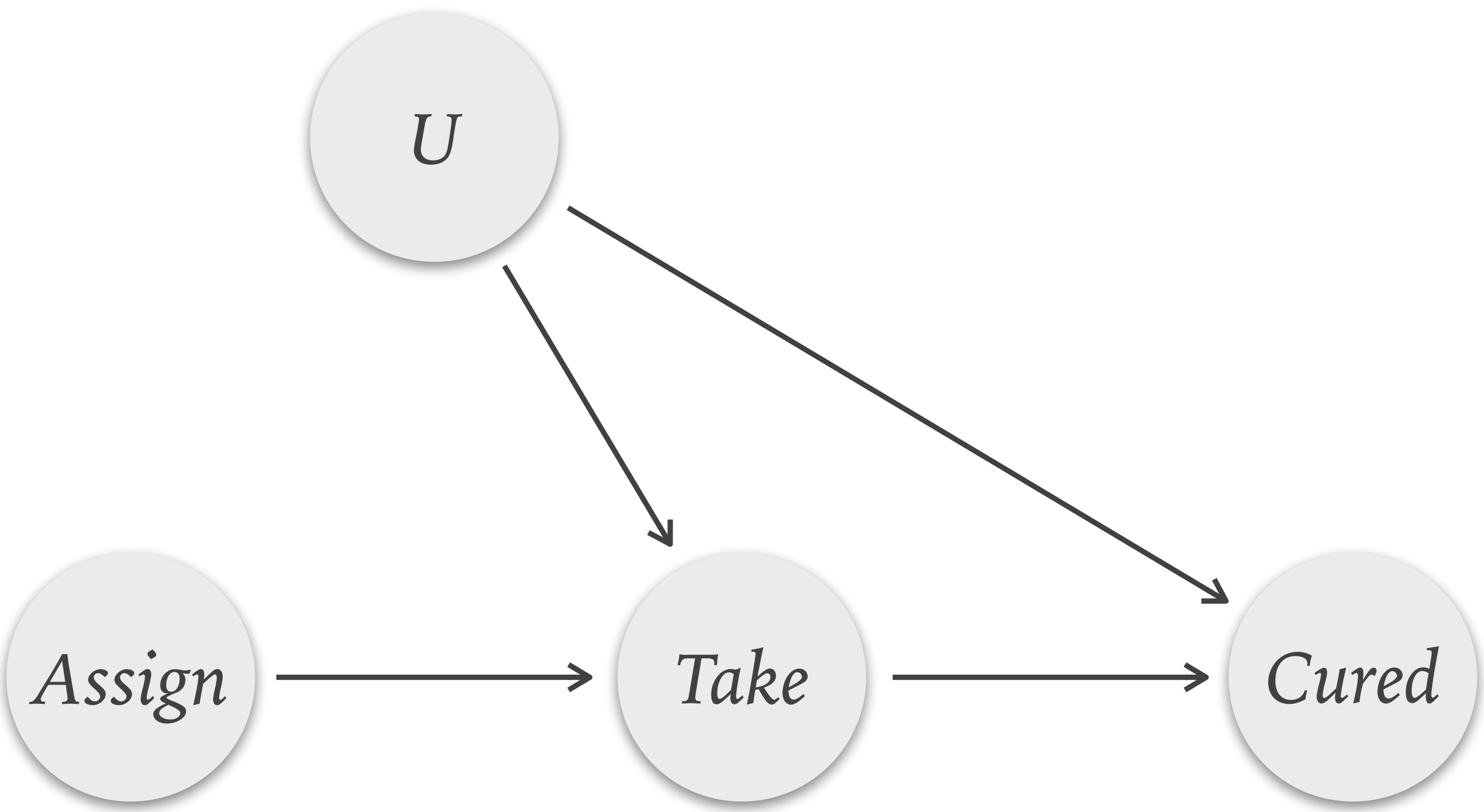
= −1, 0, 1

= 1

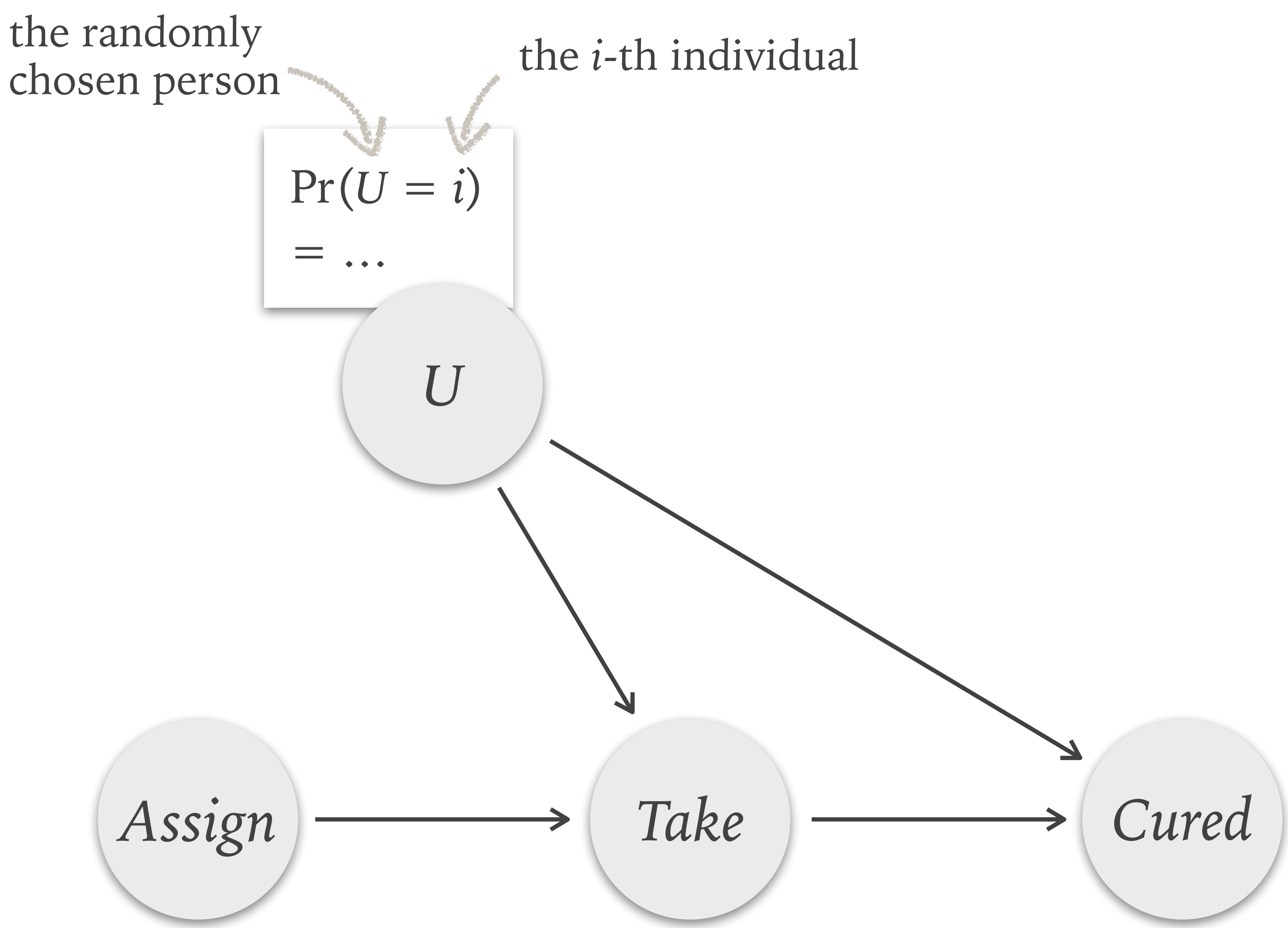
… DATE degenerates to LATE.

5. New Theorem:
An Identification Result for DATE

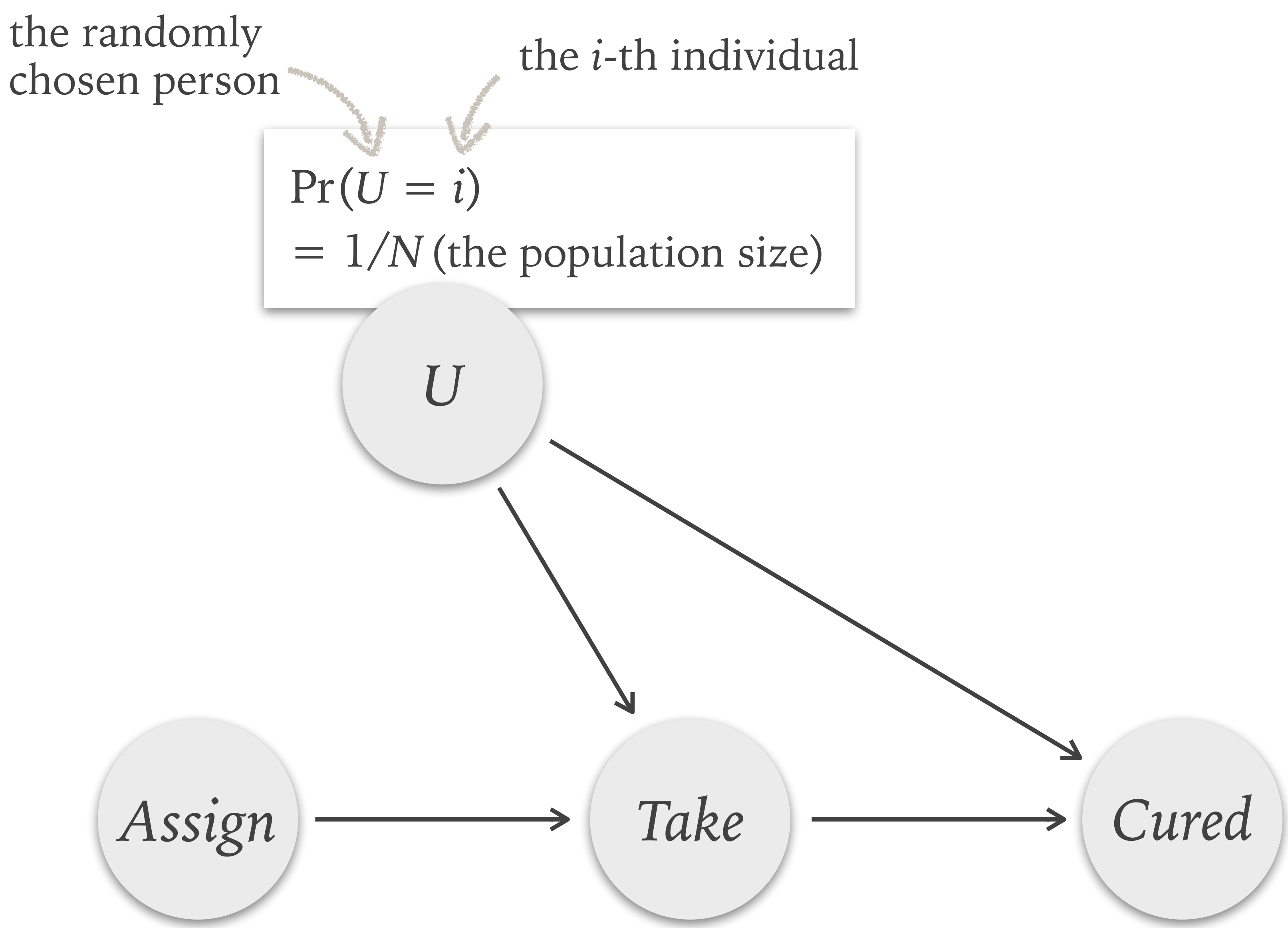
Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters



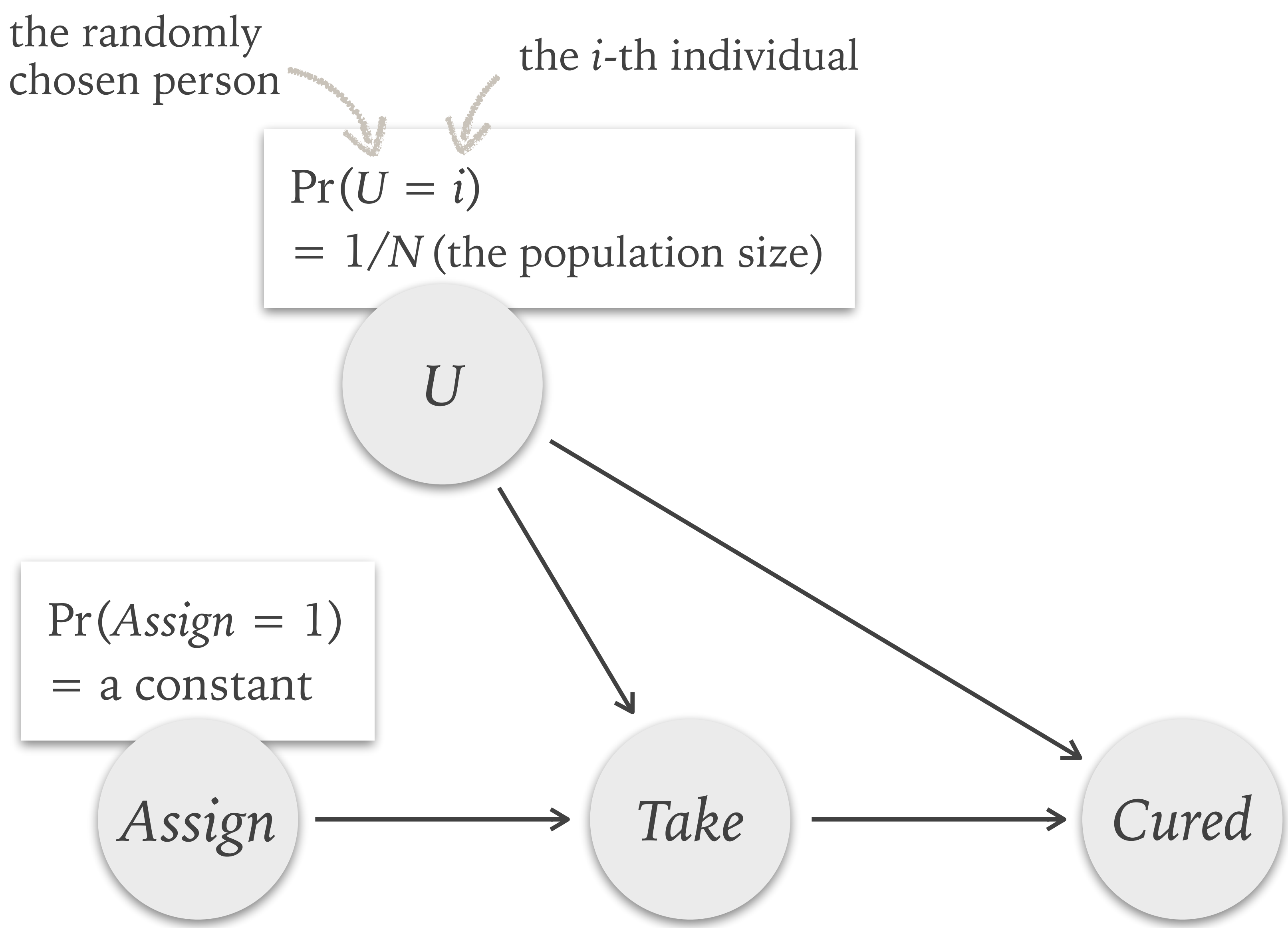
Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters



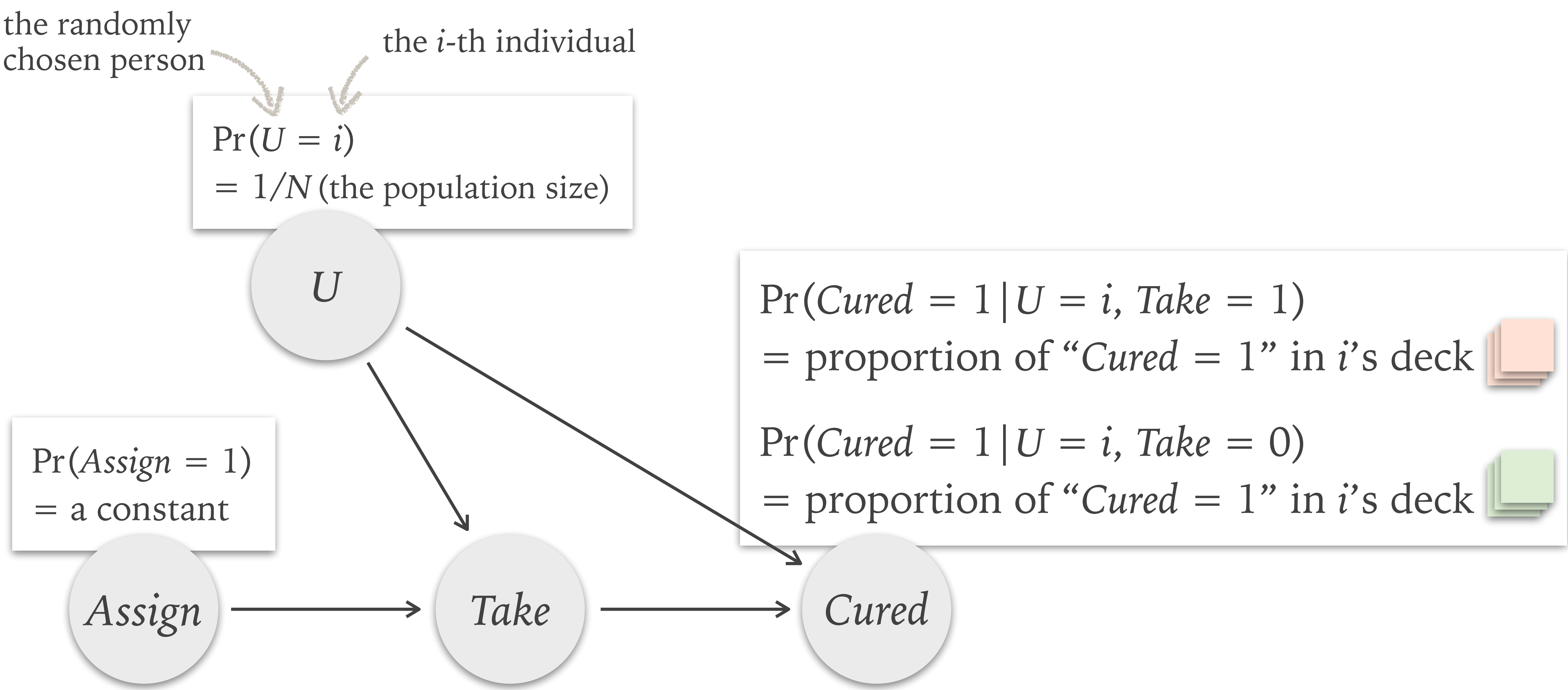
Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters



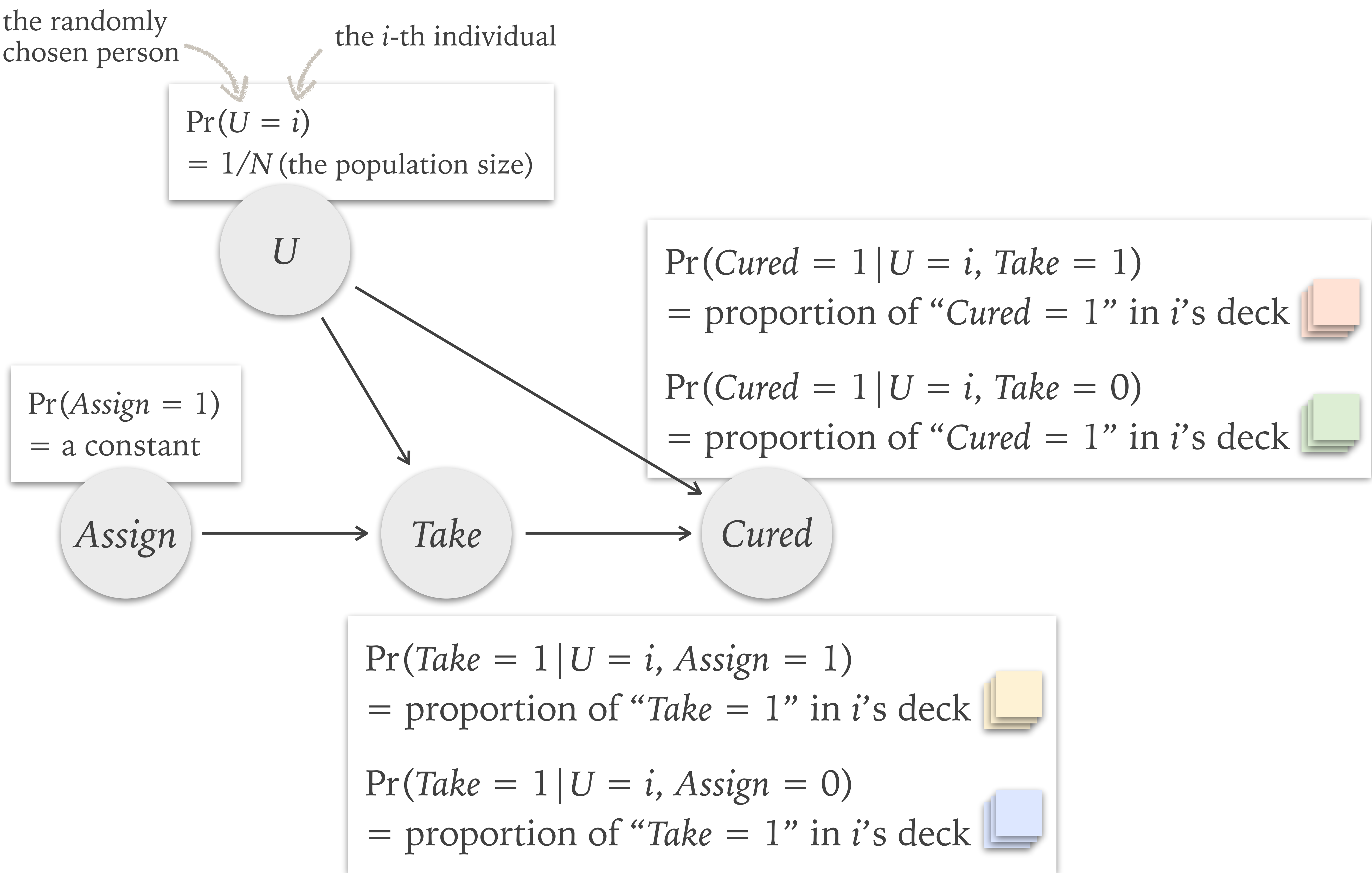
Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters



Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters

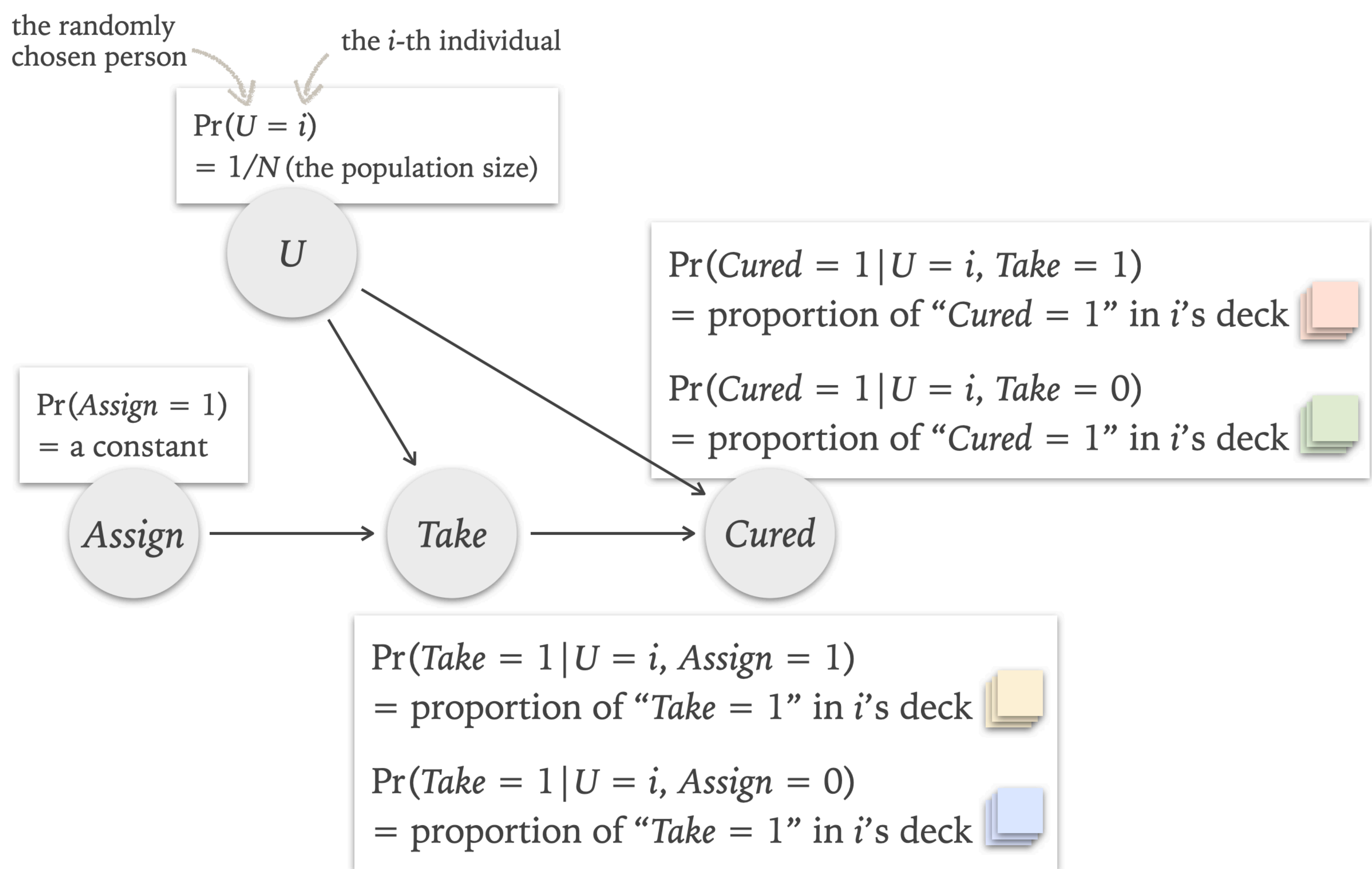


Assume the true causal model is a causal Bayes net with the following DAG and unknown parameters



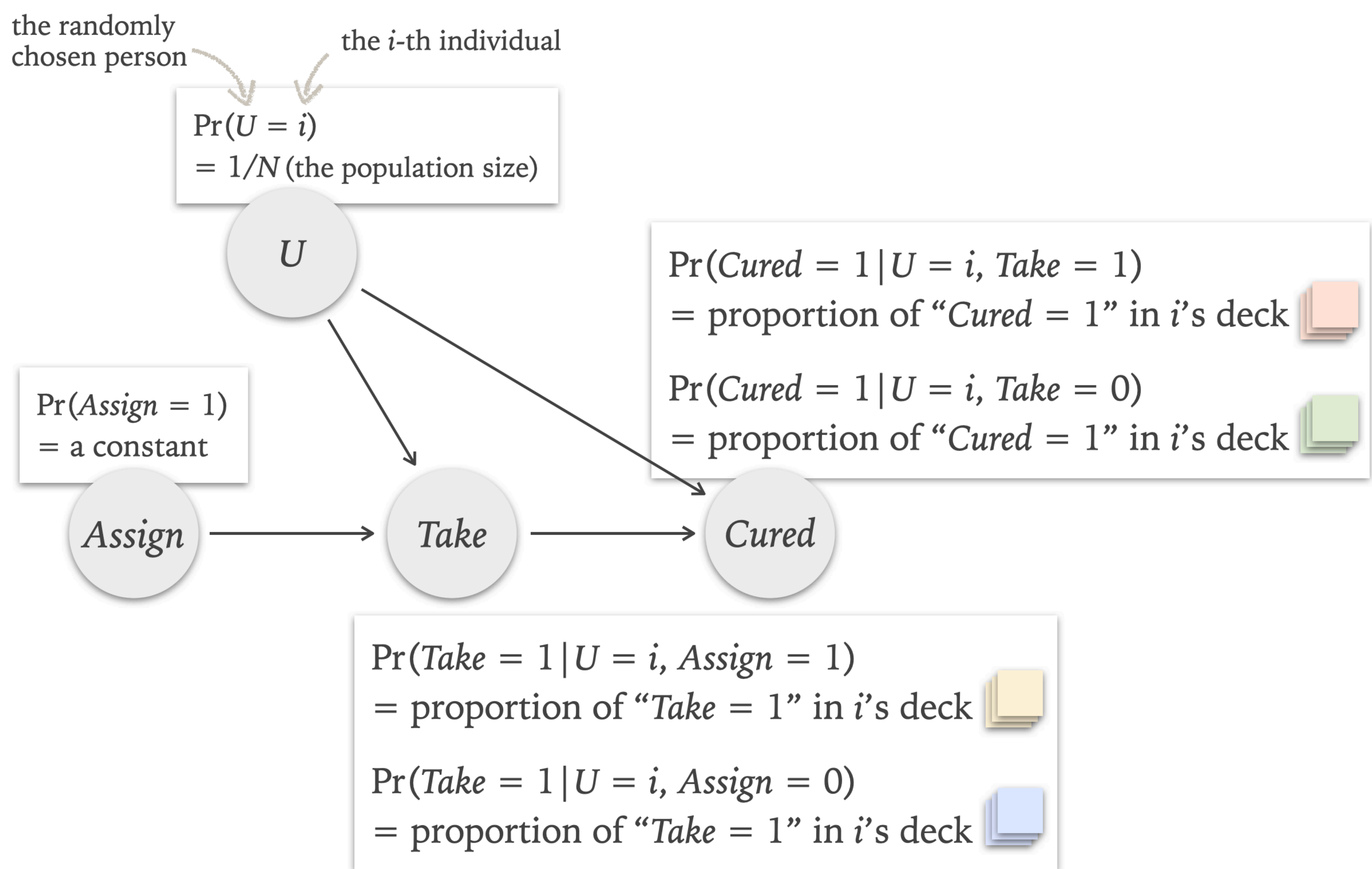
New Theorem (No Assuming CEM)

Assume that there are *no defiers* and that the true causal model is a causal Bayes net of this form:



New Theorem (No Assuming CEM)

Assume that there are *no defiers* and that the true causal model is a causal Bayes net of this form:

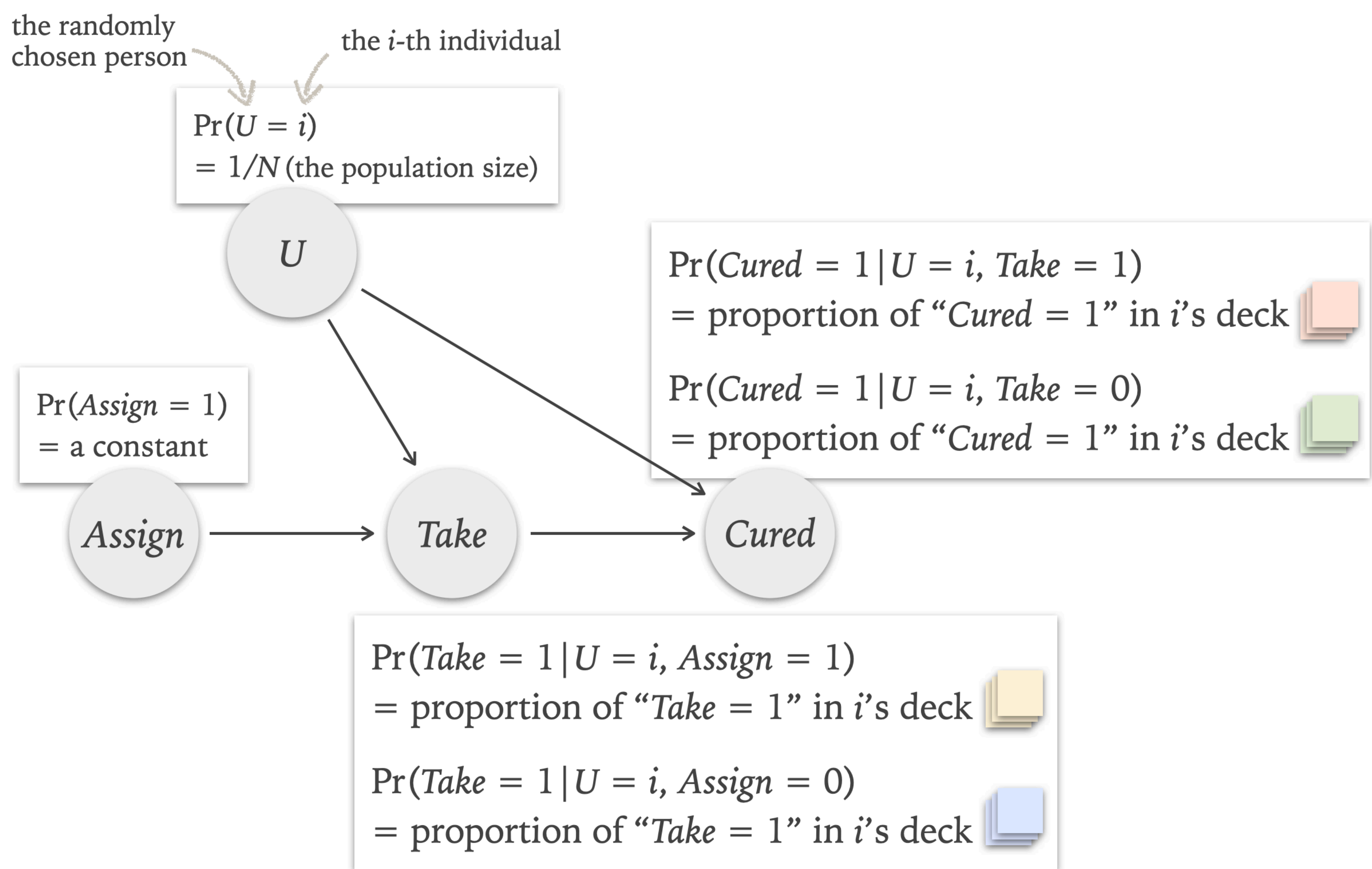


Then we have:

$$\text{DATE} = \frac{\Pr(Cured = 1 \mid Assign = 1) - \Pr(Cured = 1 \mid Assign = 0)}{\Pr(Take = 1 \mid Assign = 1) - \Pr(Take = 1 \mid Assign = 0)}$$

New Theorem (No Assuming CEM)

Assume that there are *no defiers* and that the true causal model is a causal Bayes net of this form:



Then we have:

$$\text{DATE} = \frac{\Pr(\mathit{Cured} = 1 \mid \mathit{Assign} = 1) - \Pr(\mathit{Cured} = 1 \mid \mathit{Assign} = 0)}{\Pr(\mathit{Take} = 1 \mid \mathit{Assign} = 1) - \Pr(\mathit{Take} = 1 \mid \mathit{Assign} = 0)}$$

In the special case in which every deck turns out to be a single card, this result degenerates to the classic result of LATE.

6. Wrap Up

To Dawid (2000) and other skeptics of the Rubin causal model in statistics

- I agree that CEM is invalid.
- But that poses no threat to the Rubin causal model and its application to the estimation of LATE.
- For the result of LATE can be obtained even without assuming CEM, as shown by the new theorem.

To Pearl (2009) and followers in computer science and philosophy of science

- Pearl claims that (i) structural equation models can do everything that can be done by (ii) causal Bayes nets. So, at some point, he only uses the former and no longer mentions the latter.
- I recommend a reconsideration:
  - (i) is committed to CEM, as shown by Pearl's semantics.
  - (ii) is not. This is why (ii) can do something that (i) cannot do: an identification result for DATE without assuming CEM.

To Imbens (2020) and followers in econometrics and epidemiology

- Imbens question the value of DAGs (causal graphs) in causal inference, for two reasons.
  - Not helpful for proving theorems.
  - Not helpful for stating the assumptions used in those proofs.
- I think Imbens is right when we still work with CEM.
- But things change when we wish to drop CEM.
  - When we try to give a fully stochastic update to the Rubin causal model and the result of LATE, it is easy to do it with a causal Bayes net and the DAG that comes with it.
- Let me elaborate on the next page …

Potential Outcomes

ordinary potential outcome

$$Cured_i^{t=1}$$

= the medical result (being cured, or not) that i would have if i took the treatment

= Boolean-valued, 0 or 1 (in the Rubin causal model)

Probabilistic Potential Outcomes

ordinary potential outcome

$Cured_i^{t=1}$

= the medical result (being cured, or not) that i would have if i took the treatment

= Boolean-valued, 0 or 1 (in the Rubin causal model)

↓ a stochastic upgrade by Robins & Greenland (1989, 2000) on "probability of causation", but sometimes not easy to use to state assumptions

probabilistic version

$\theta_i^{t=1}$

Probabilistic Potential Outcomes

ordinary potential outcome

$Cured_i^{t=1}$

= the medical result (being cured, or not) that i would have if i took the treatment

= Boolean-valued, 0 or 1 (in the Rubin causal model)

↓ **a stochastic upgrade by Robins & Greenland (1989, 2000) on "probability of causation", but sometimes not easy to use to state assumptions**

probabilistic version

$\theta_i^{t=1}$

= the probability of being cured that i would have if i took the treatment

= the proportion of "$Cured = 1$" cards in a certain deck that i possesses

= probability-valued, an unknown parameter in [0, 1]

written $\Pr_i^{t=1}(Cured = 1)$ in the above

Probabilistic Potential Outcomes as Parameters of a Causal Bayes Net

ordinary potential outcome

$Cured_i^{t=1}$

= the medical result (being cured, or not) that i would have if i took the treatment

= Boolean-valued, 0 or 1 (in the Rubin causal model)

a stochastic upgrade by Robins & Greenland (1989, 2000) on "probability of causation", but sometimes not easy to use to state assumptions

probabilistic version

$\theta_i^{t=1}$

= the probability of being cured that i would have if i took the treatment

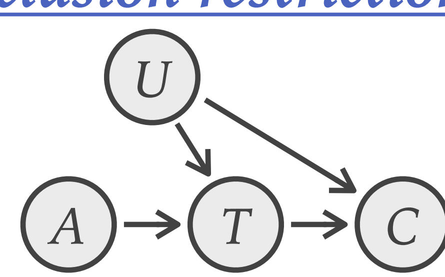
= the proportion of "$Cured = 1$" cards in a certain deck that i possesses

= probability-valued, an unknown parameter in [0, 1]

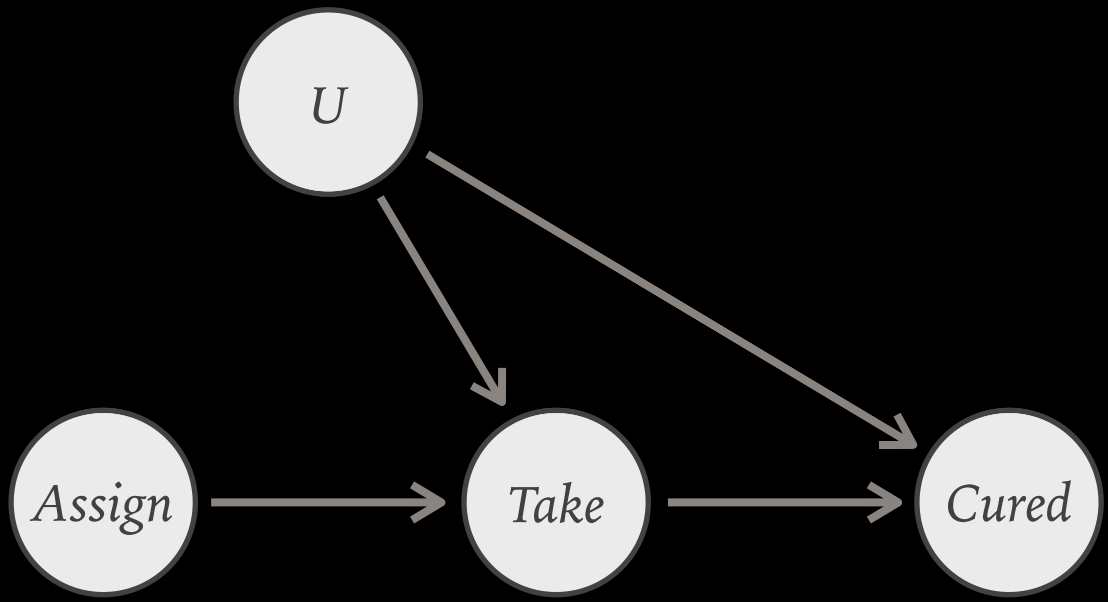
written $\Pr_i^{t=1}(Cured = 1)$ in the above

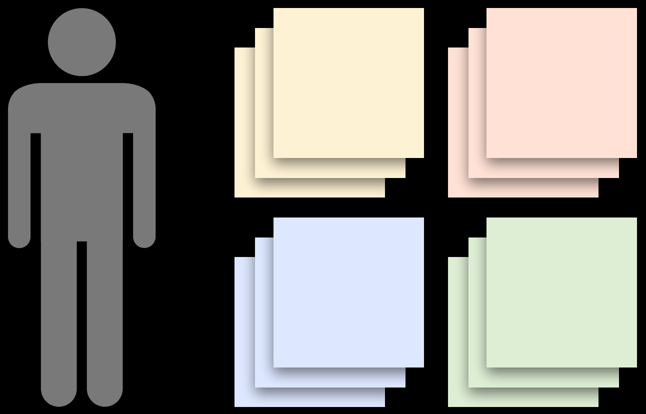
a parameter in a causal Bayes net

Then we can use a DAG to easily express assumptions (such as *exclusion restriction*). The proof for DATE is simple, too.



Thank You!





Proof

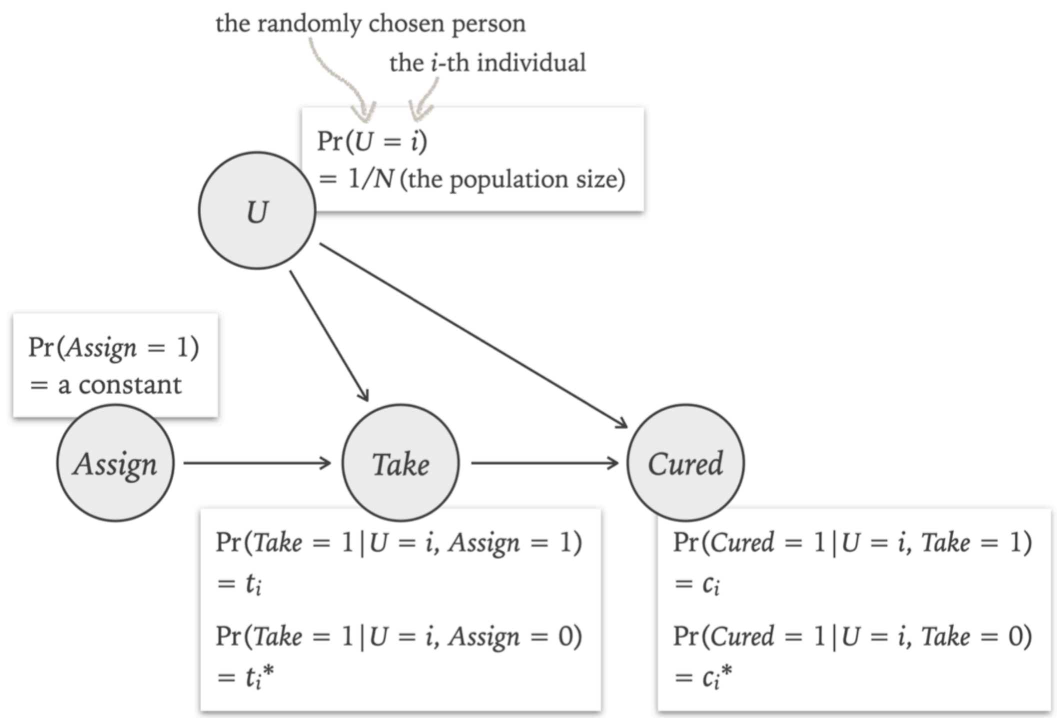


Figure 2: The causal Bayes net assumed in Theorem 2

(direct causes). Start with the first term in the numerator:

$$
\begin{aligned}
&\Pr(\mathit{Cured} = 1 \mid \mathit{Assign} = 1) \\
&= \sum_{i,j} \Big(\Pr(\mathit{Cured} = 1 \mid \mathit{Take} = j,\, U = i,\, \mathit{Assign} = 1) \\
&\qquad \times \Pr(\mathit{Take} = j \mid U = i,\, \mathit{Assign} = 1) \\
&\qquad \times \Pr(U = i \mid \mathit{Assign} = 1) \Big) && \text{by Chain Rule} \\
&= \sum_{i,j} \Big(\Pr(\mathit{Cured} = 1 \mid \mathit{Take} = j,\, U = i,\, \cancel{\mathit{Assign} = 1}) \\
&\qquad \times \Pr(\mathit{Take} = j \mid U = i,\, \mathit{Assign} = 1) \\
&\qquad \times \Pr(U = i \mid \cancel{\mathit{Assign} = 1}) \Big) && \text{by Causal Markov} \\
&= \sum_{i} \Big(\Pr(\mathit{Cured} = 1 \mid \mathit{Take} = 1,\, U = i) \\
&\qquad \times \Pr(\mathit{Take} = 1 \mid U = i,\, \mathit{Assign} = 1) \\
&\qquad \times \Pr(U = i) \Big) \\
&\quad + \sum_{i} \Big(\Pr(\mathit{Cured} = 1 \mid \mathit{Take} = 0,\, U = i) \\
&\qquad \times \Pr(\mathit{Take} = 0 \mid U = i,\, \mathit{Assign} = 1) \\
&\qquad \times \Pr(U = i) \Big) \\
&= \sum_i \Big(c_i\, t_i\, \tfrac{1}{N} \Big) + \sum_i \Big(c_i^*\, (1 - t_i)\, \tfrac{1}{N} \Big) \\
&= \tfrac{1}{N} \sum_i \Big(c_i\, t_i + c_i^*\, (1 - t_i) \Big).
\end{aligned}
$$

Similarly for the second term in the numerator:

$$
\begin{aligned}
&\Pr(\mathit{Cured} = 1 \mid \mathit{Assign} = 0) \\
&= \tfrac{1}{N} \sum_i \Big(c_i\, t_i^* + c_i^*\, (1 - t_i^*) \Big).
\end{aligned}
$$

Proof

Now calculate the first term in the denominator:

$$
\begin{aligned}
&\Pr(\mathit{Take}=1 \,|\, \mathit{Assign}=1) \\
&= \sum_i \Big(\Pr(\mathit{Take}=1 \,|\, U=i, \mathit{Assign}=1) \cdot \Pr(U=i \,|\, \underbrace{\cancel{\mathit{Assign}=1}}_{\text{by Causal Markov}}) \Big) \\
&= \sum_i t_i \tfrac{1}{N} \\
&= \tfrac{1}{N} \sum_i t_i \,.
\end{aligned}
$$

Similarly for the second term in the denominator:

$$
\begin{aligned}
&\Pr(\mathit{Take}=1 \,|\, \mathit{Assign}=0) \\
&= \tfrac{1}{N} \sum_i t_i^* \,.
\end{aligned}
$$

To finish off, plug the four terms just calculated into the following:

$$
\begin{aligned}
&\frac{\Pr(\mathit{Cured}=1 \,|\, \mathit{Assign}=1) \;-\; \Pr(\mathit{Cured}=1 \,|\, \mathit{Assign}=0)}{\Pr(\mathit{Take}=1 \,|\, \mathit{Assign}=1) \;-\; \Pr(\mathit{Take}=1 \,|\, \mathit{Assign}=0)} \\
&= \frac{\cancel{\tfrac{1}{N}} \sum_i \big(c_i\, t_i + c_i^*\,(1-t_i)\big) - \cancel{\tfrac{1}{N}} \sum_i \big(c_i\, t_i^* + c_i^*\,(1-t_i^*)\big)}{\cancel{\tfrac{1}{N}} \sum_i t_i - \cancel{\tfrac{1}{N}} \sum_i t_i^*} \\
&= \frac{\sum_i (t_i - t_i^*)(c_i - c_i^*)}{\sum_j (t_j - t_j^*)} \\
&= \sum_i \left(\frac{t_i - t_i^*}{\sum_j (t_j - t_j^*)} \right) (c_i - c_i^*) \\
&= \mathrm{DATE}\,,
\end{aligned}
$$

as desired. □