

Challenging the theory-based/ counterfactual binary

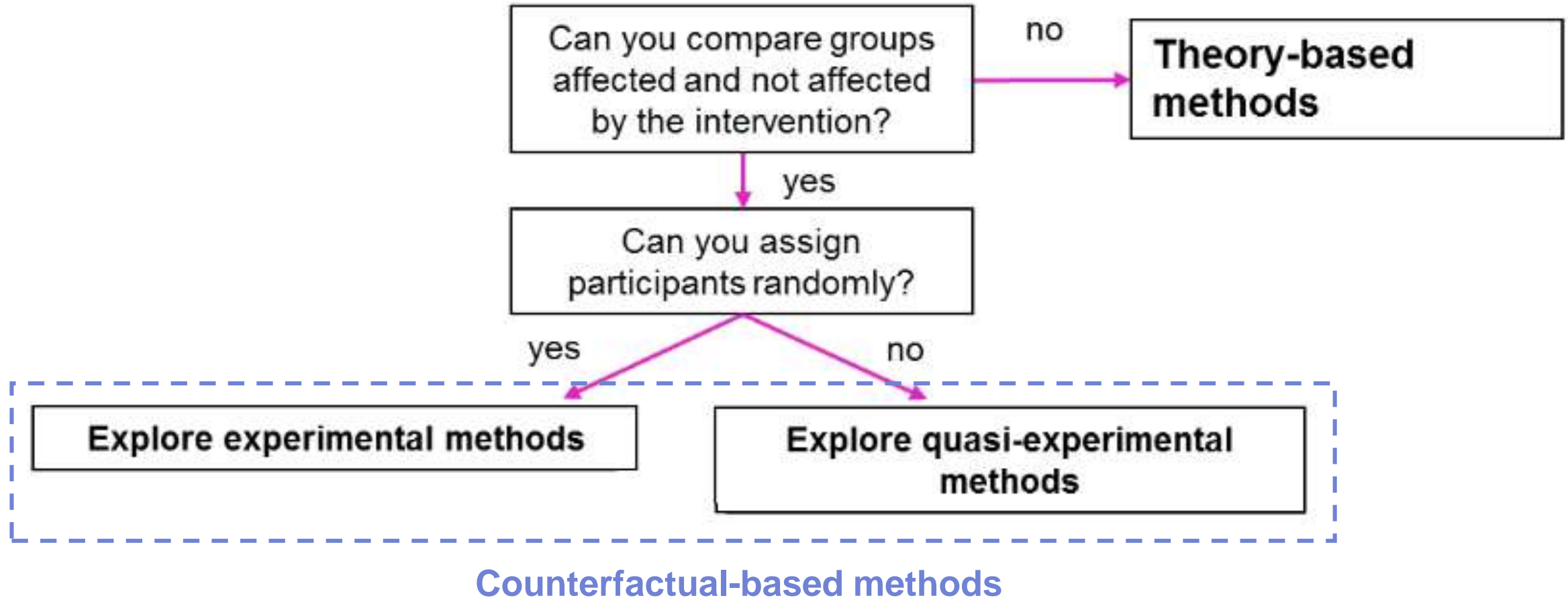
Dr Andi Fugard

Co-Director, NatCen's Centre for Evaluation

 andi.fugard@natcen.ac.uk



The Magenta book (2020)



The plan

In what sense might an RCT or quasi-experiment be “theory-based”?

Can a counterfactual-based evaluation be conducted without a comparison group?

To discuss: implications for evaluation practice

**In what sense
might an RCT or
quasi-experiment
be “theory-based”?**

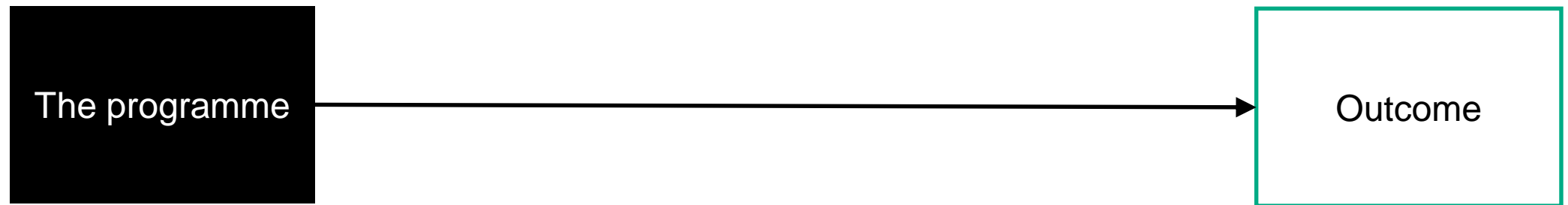


A basic problem that requires theory: selecting variables



The programme

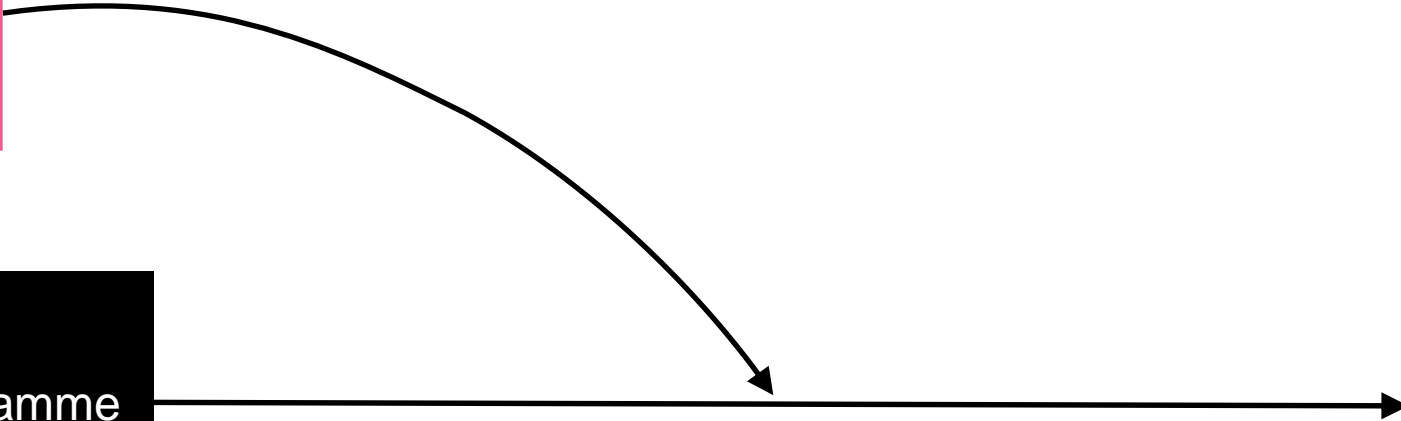
A basic problem that requires theory: selecting variables



Moderator

The programme

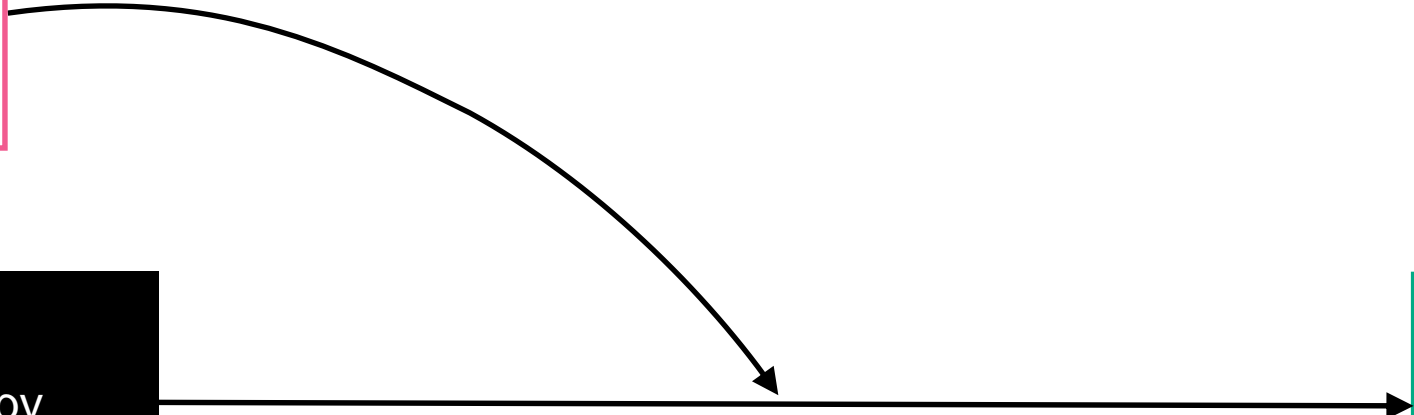
Outcome



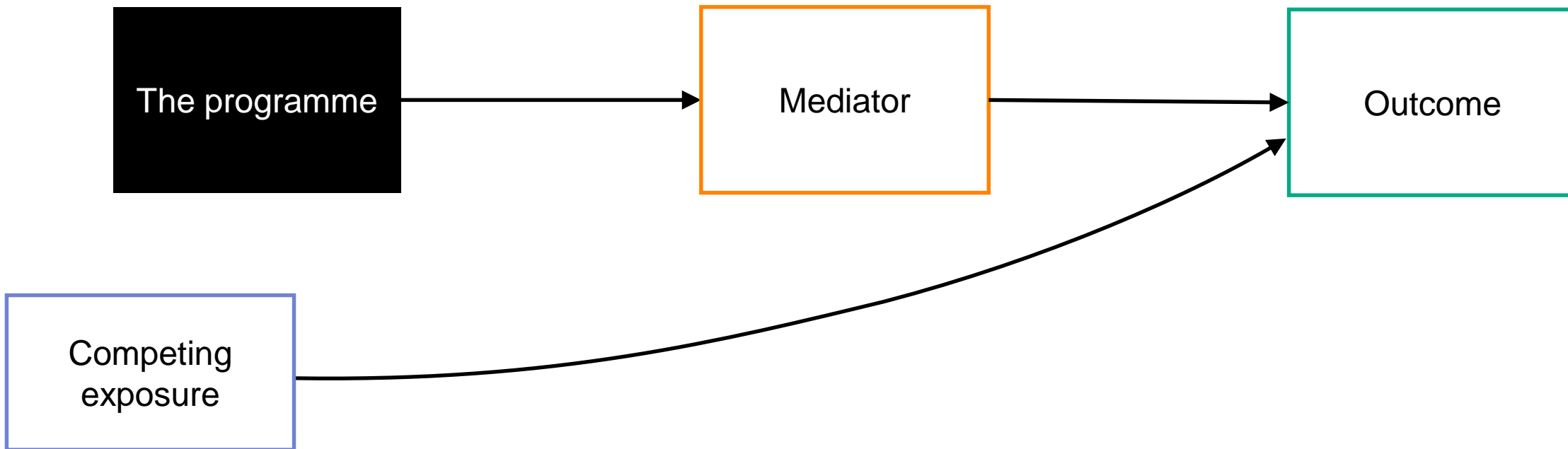
Therapist
competence

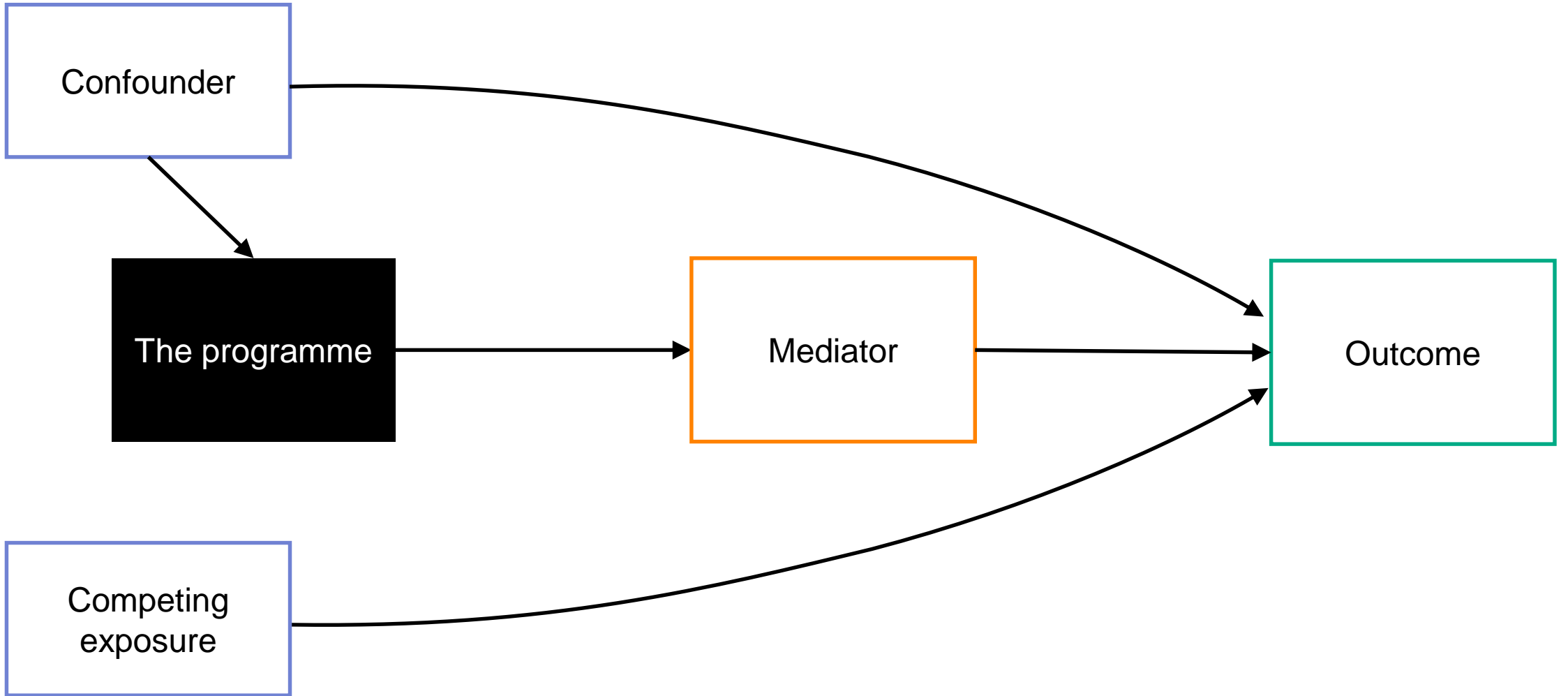
Therapy

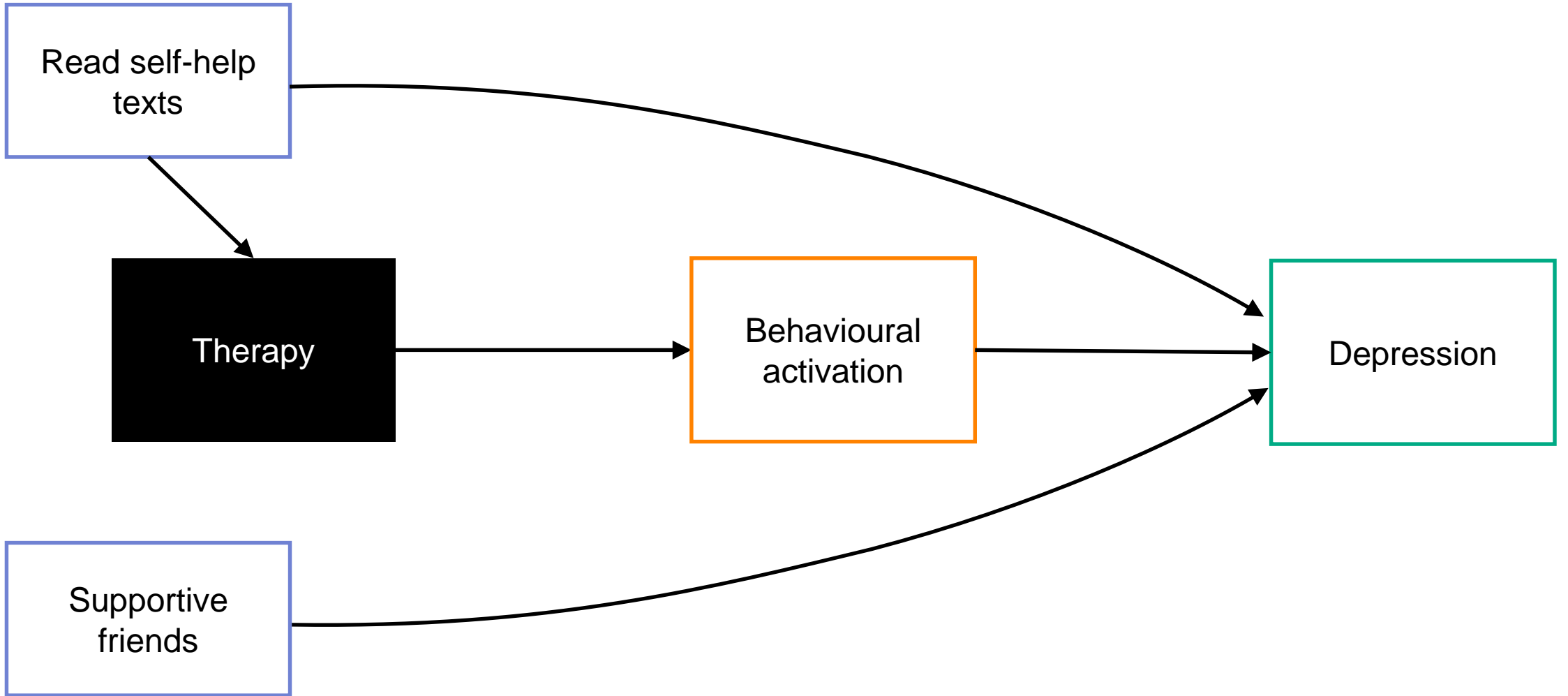
Depression

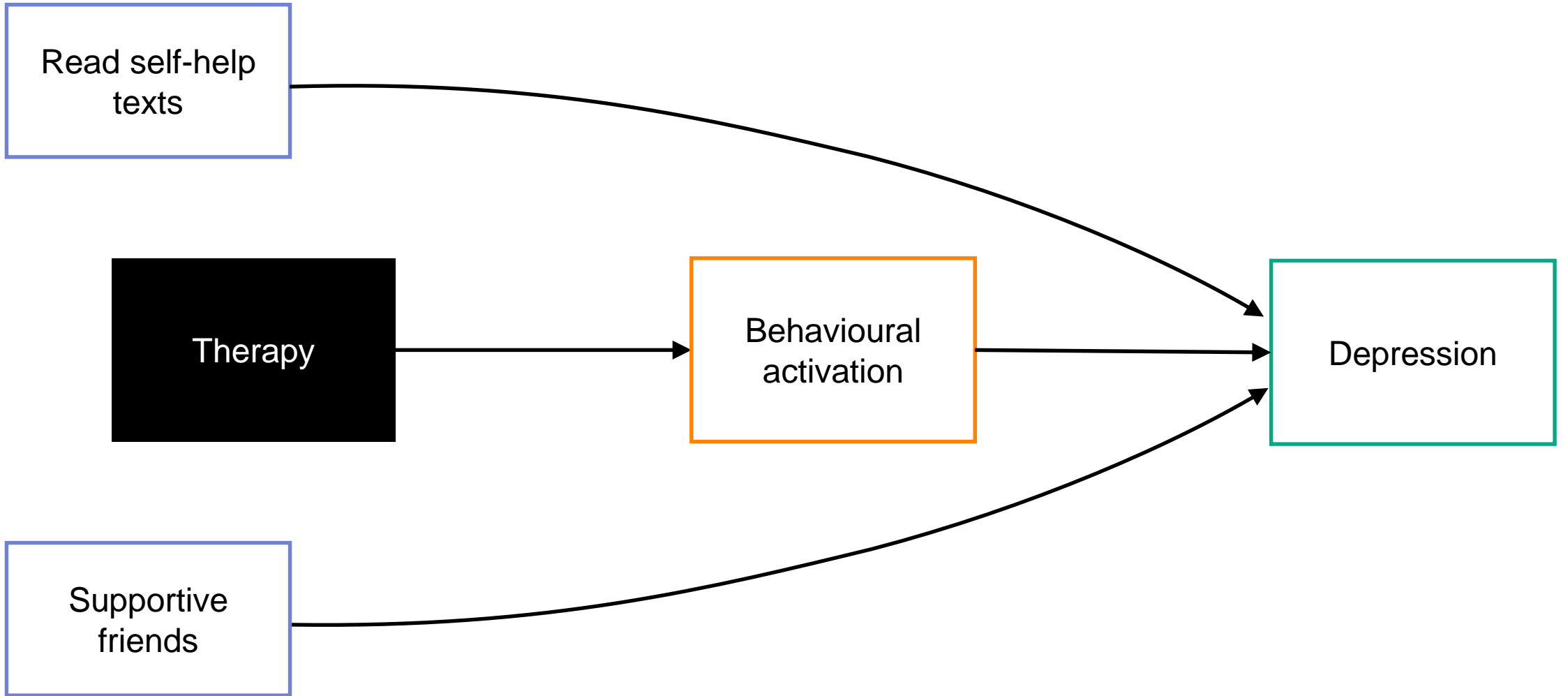












The first definition of something named “theory-based evaluation”

“A theory-based evaluation of a program is one in which the selection of program features to evaluate is determined by an explicit conceptualization of the program in terms of a theory [...] which attempts to explain how the program produces the desired effects. The theory might be psychological [...] or social psychological [...] or philosophical [...]. The essential characteristic is that the theory points out a causal relationship between a process *A* and an outcome *B*.”

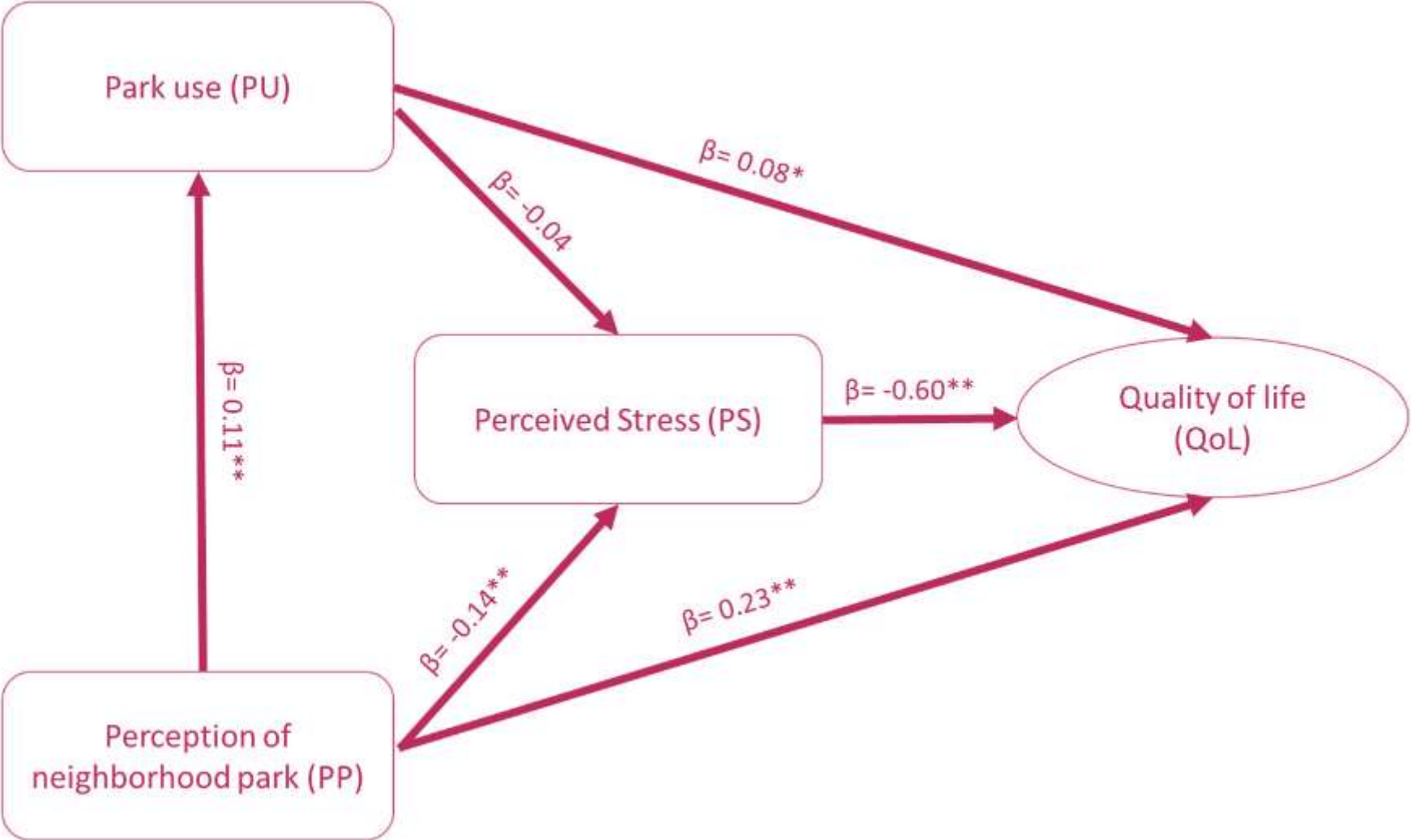
Fitz-Gibbon, C. T., & Morris, L. L. (1975).
Theory-based evaluation.
Evaluation Comment, 5(1), 1–4

Theory-based evaluation has included statistical models

“The basic task is to see how well the evidence matches the theories that were posited. Path analysis is conceptually compatible with TBE and has been used by evaluators [...].”

Carol Weiss (1997, p. 512). How can theory-based evaluation make greater headway? *Evaluation Review*, 21(4), 501–524.

Path analysis / structural equation model



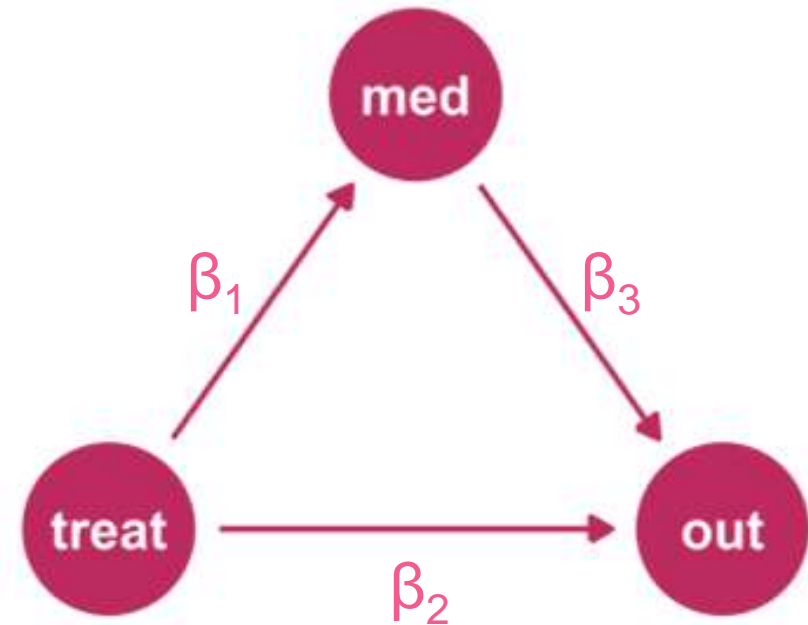
Another basic problem that requires theory: defining the causal structure

Suppose we have *all*
the variables we need...

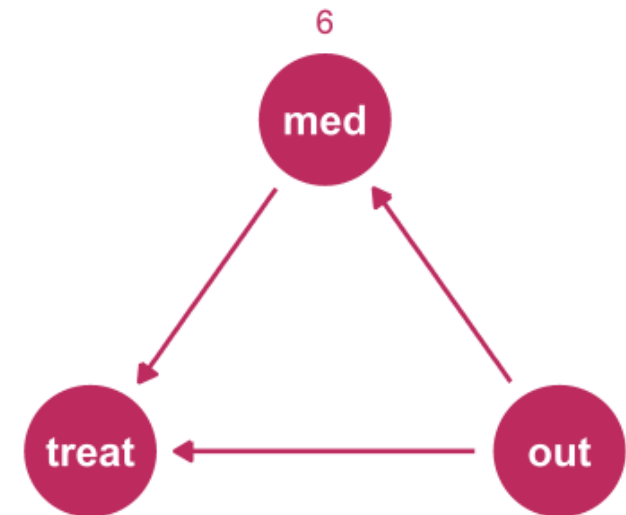
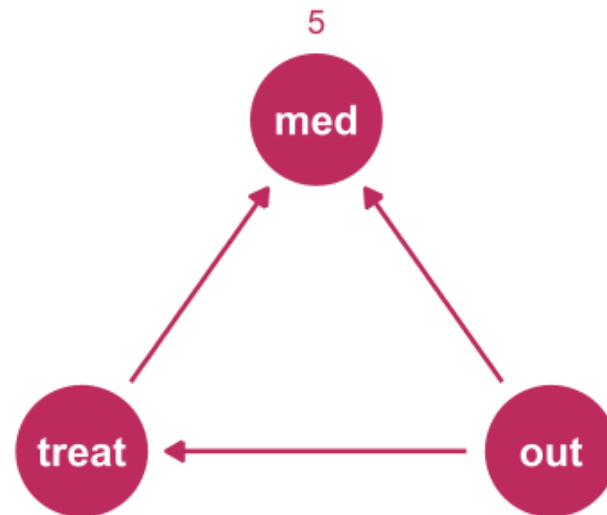
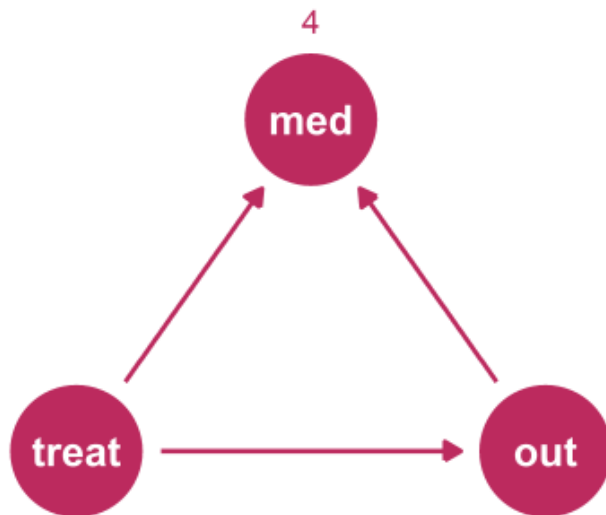
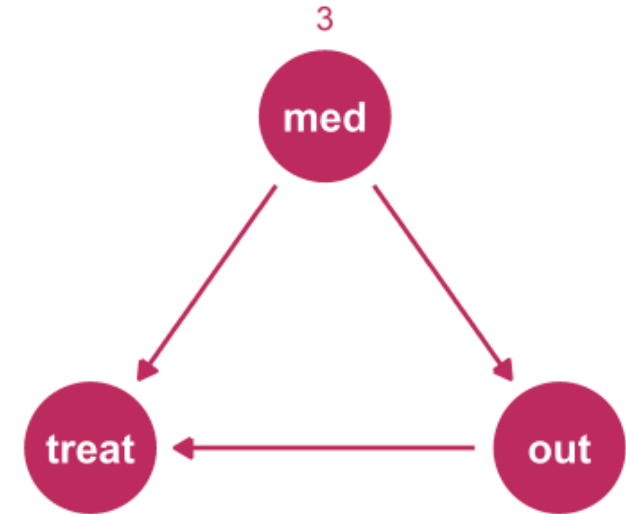
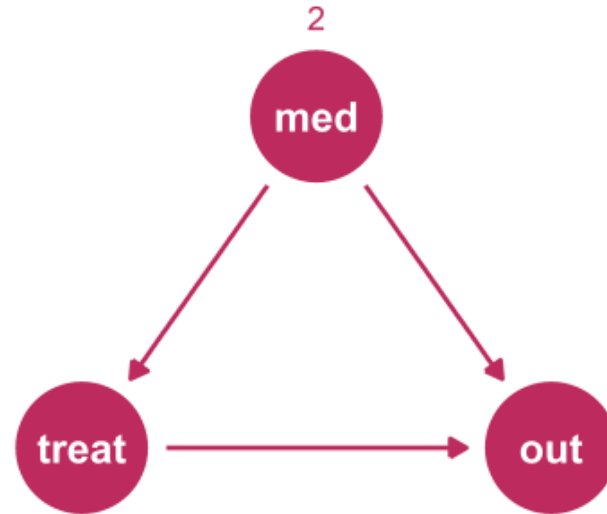
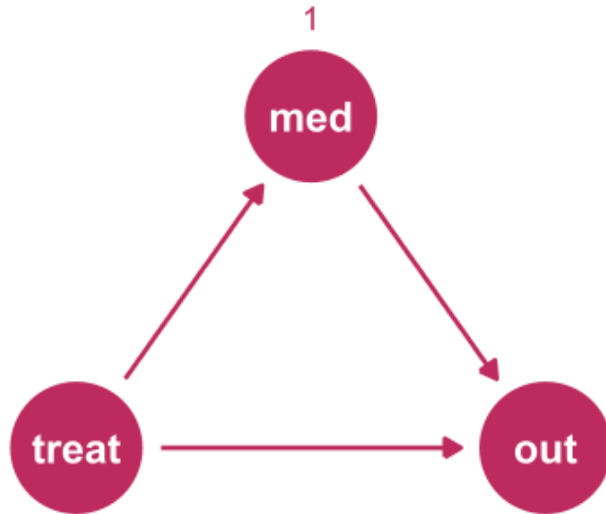
... and find these
statistical associations:

$$med = \beta_1 treat$$

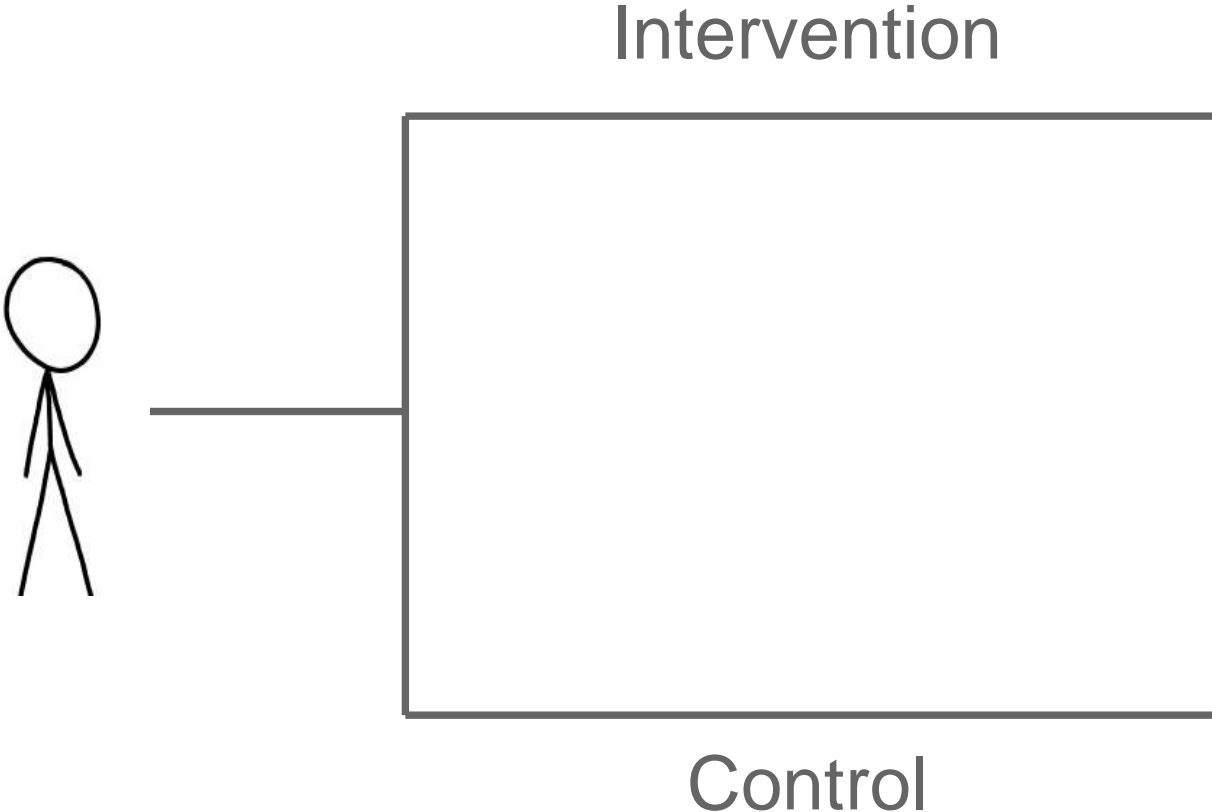
$$out = \beta_2 treat + \beta_3 med$$



All the following causal structures are *Markov equivalent*

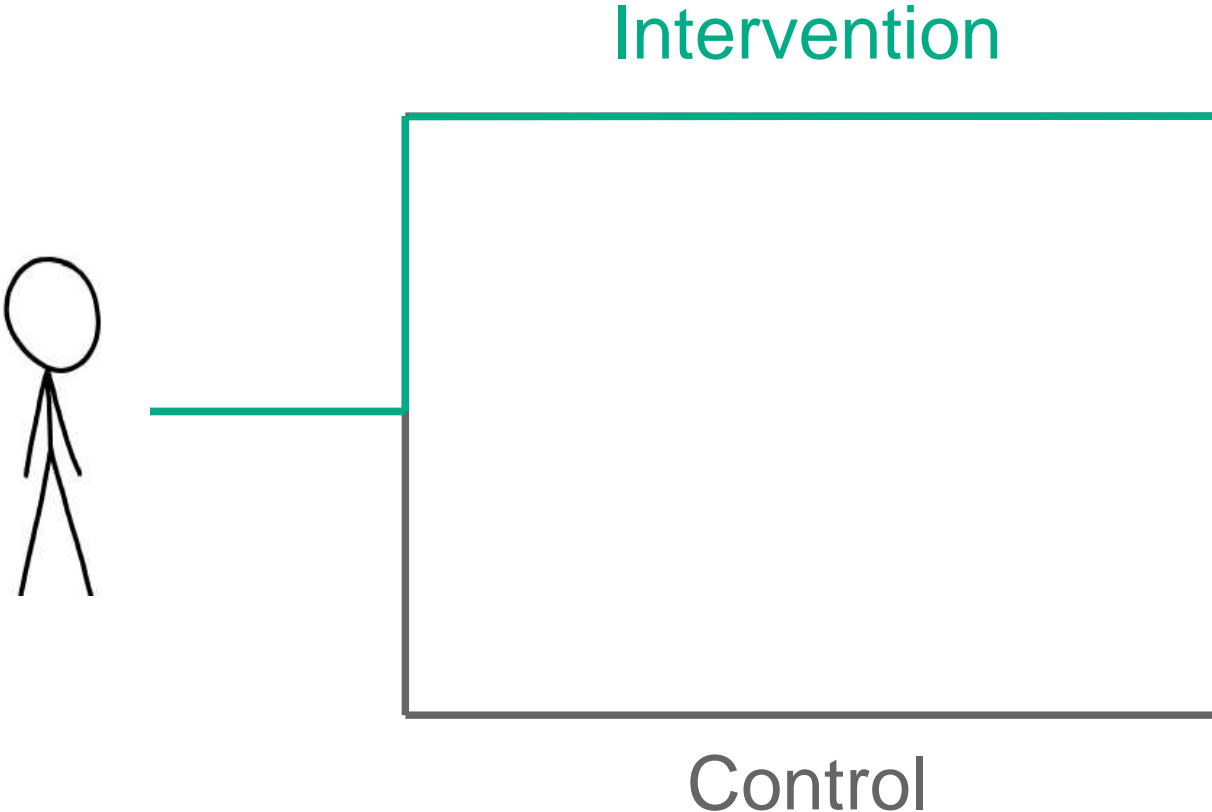


Alex enters a randomised controlled trial

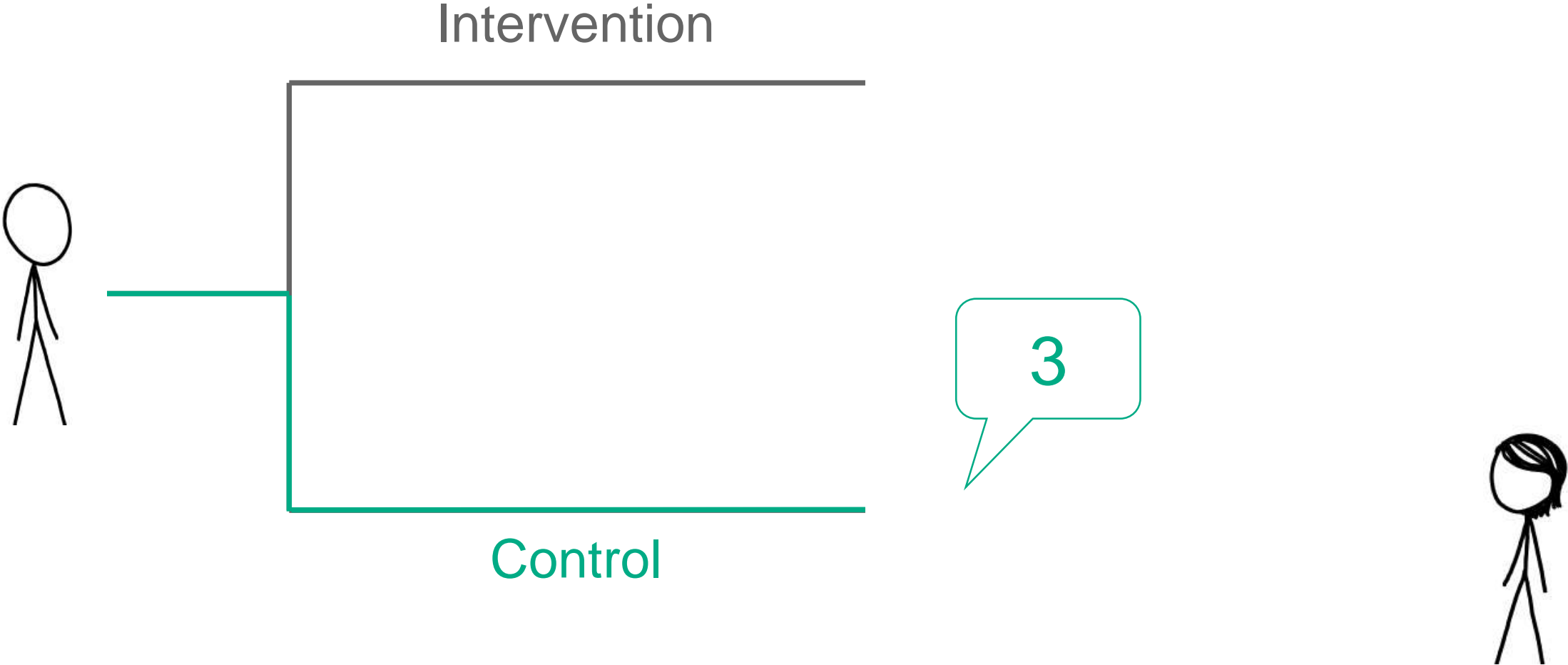


Randomised to intervention

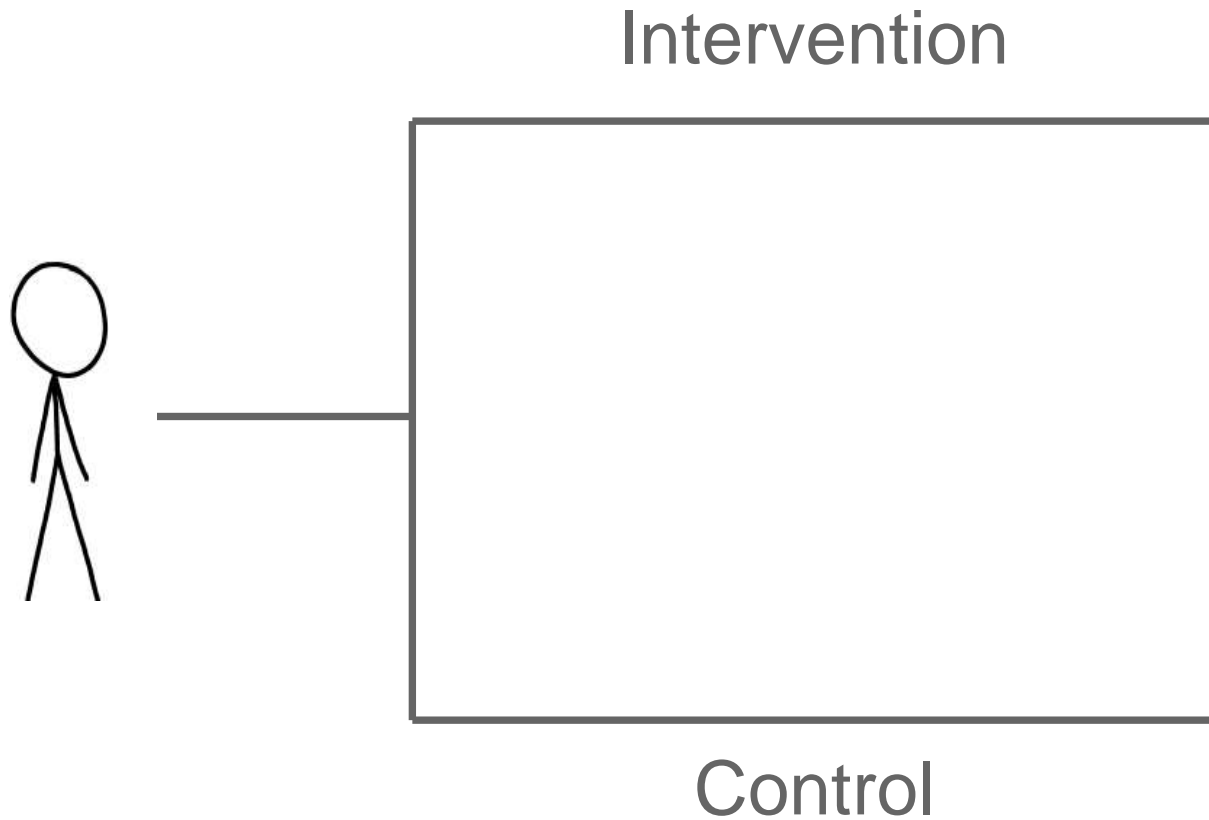
8



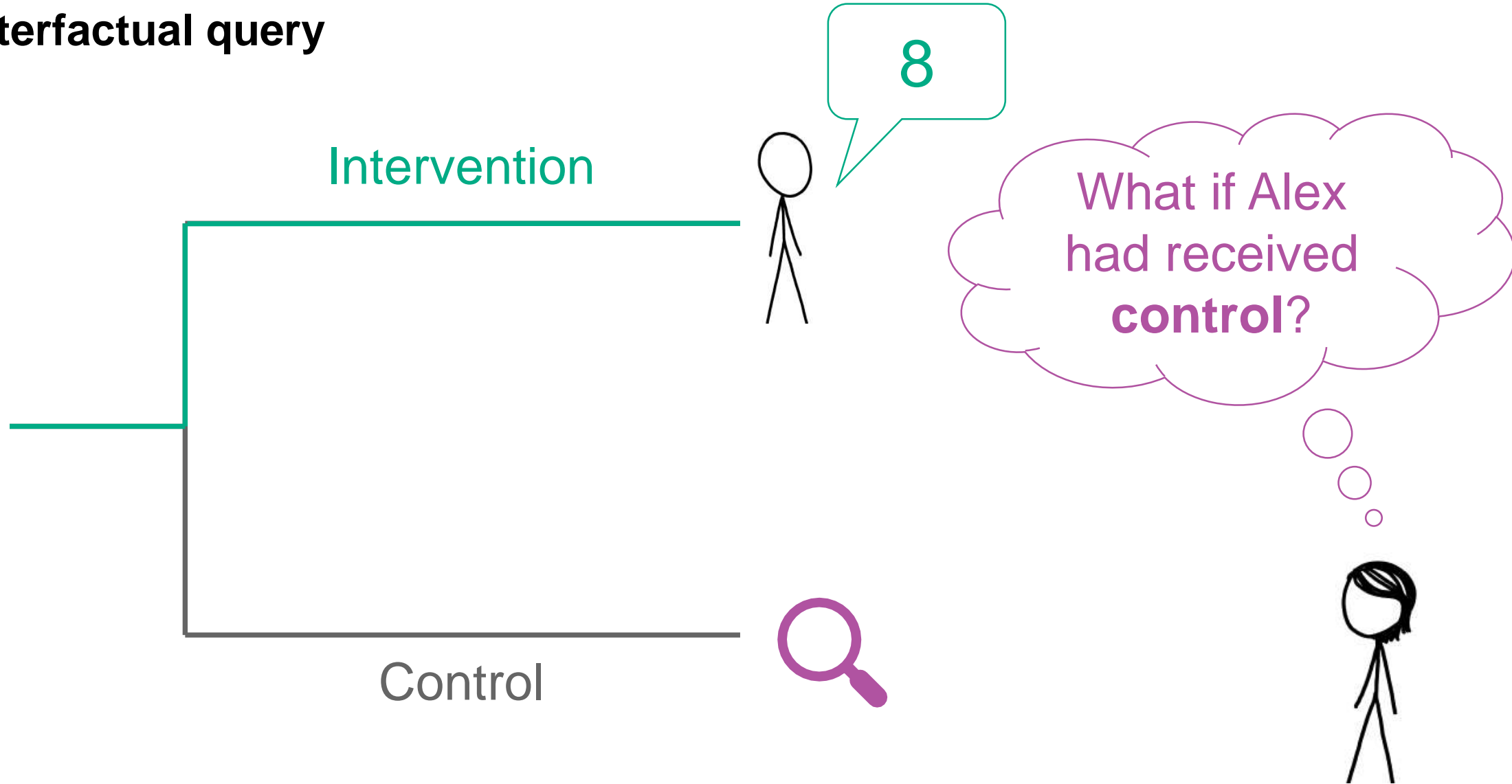
Randomised to control



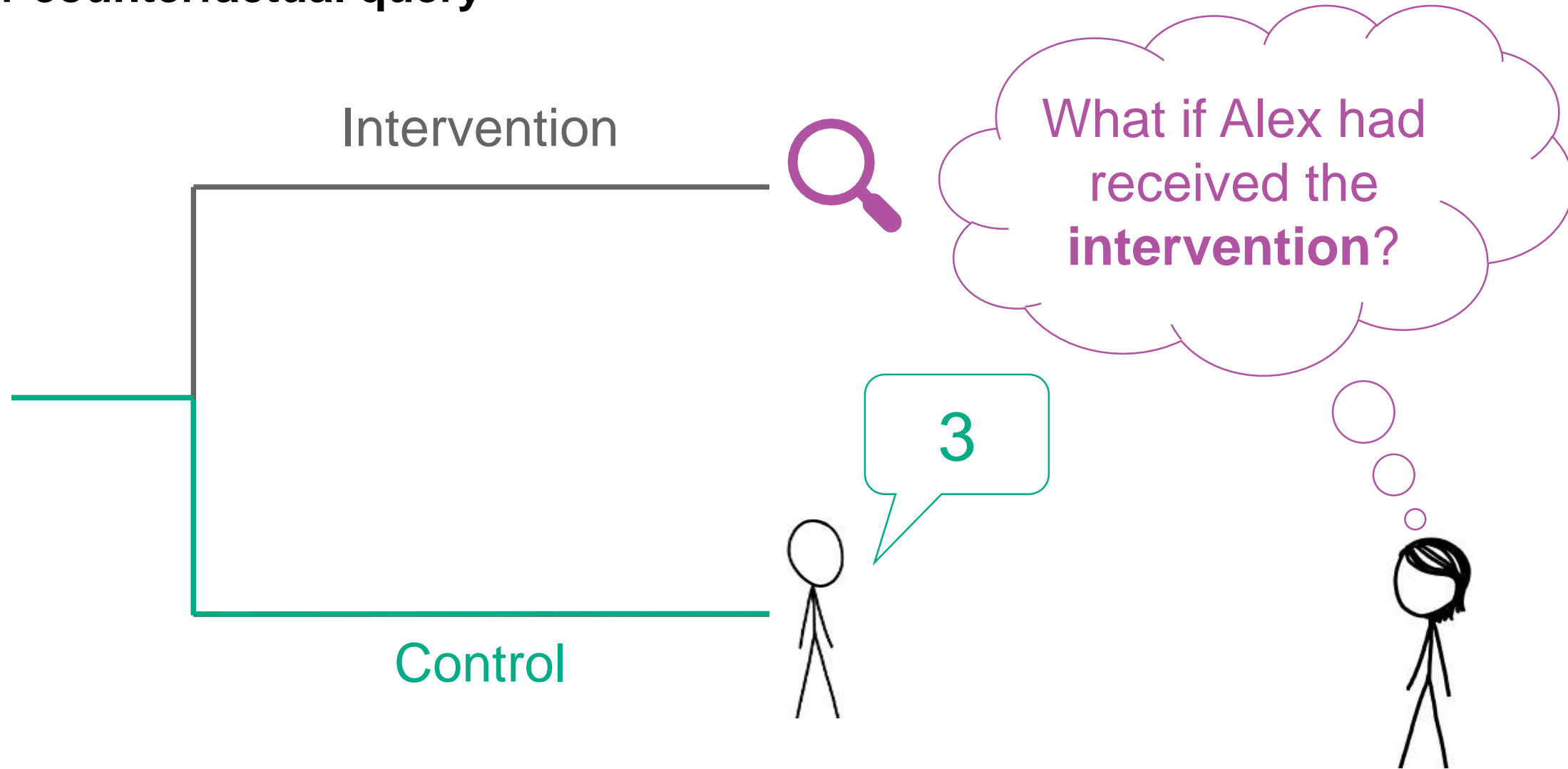
Where is the counterfactual...?



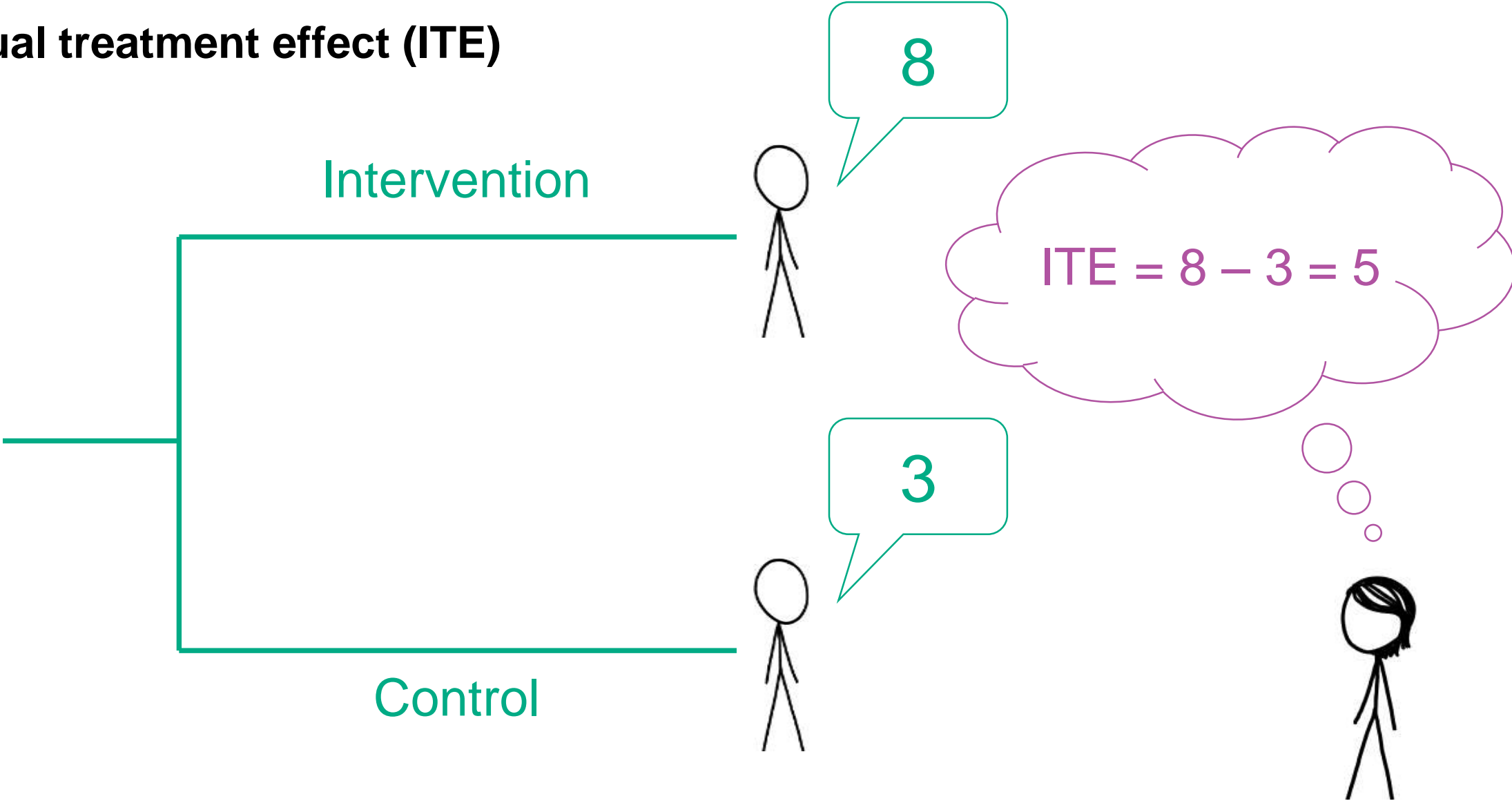
A counterfactual query



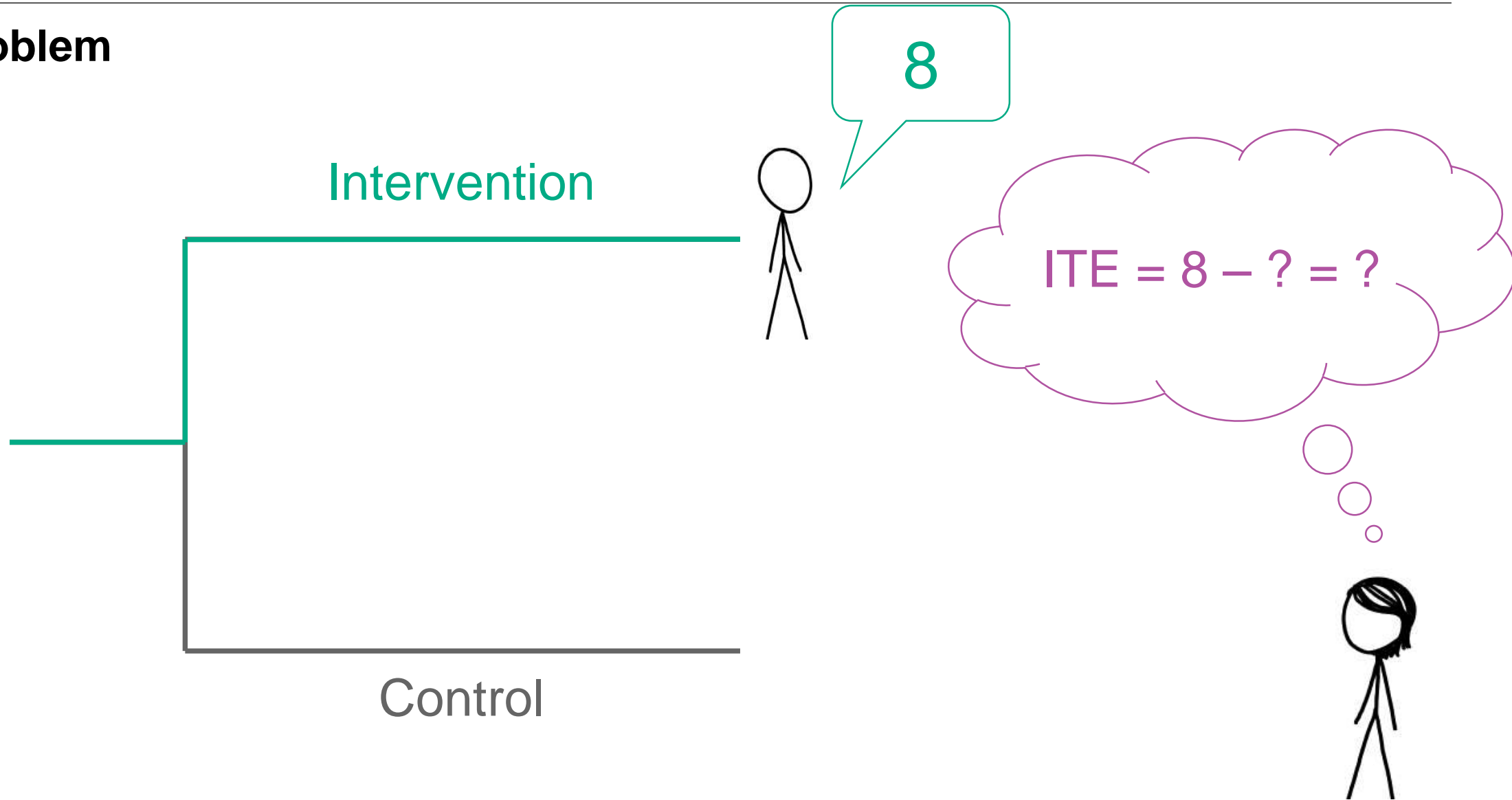
Another counterfactual query



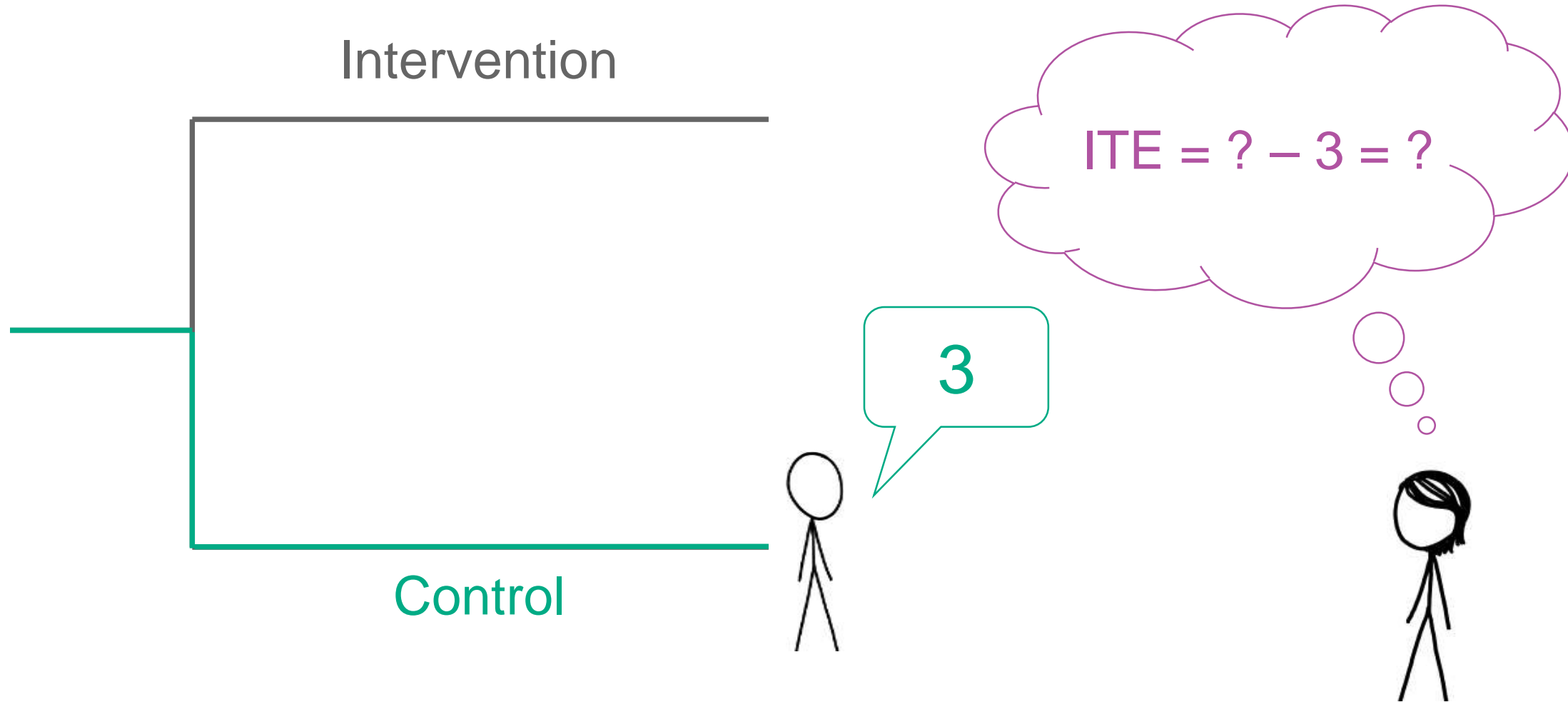
Individual treatment effect (ITE)



The problem



The problem



The fundamental problem of causal inference and how an RCT solves it

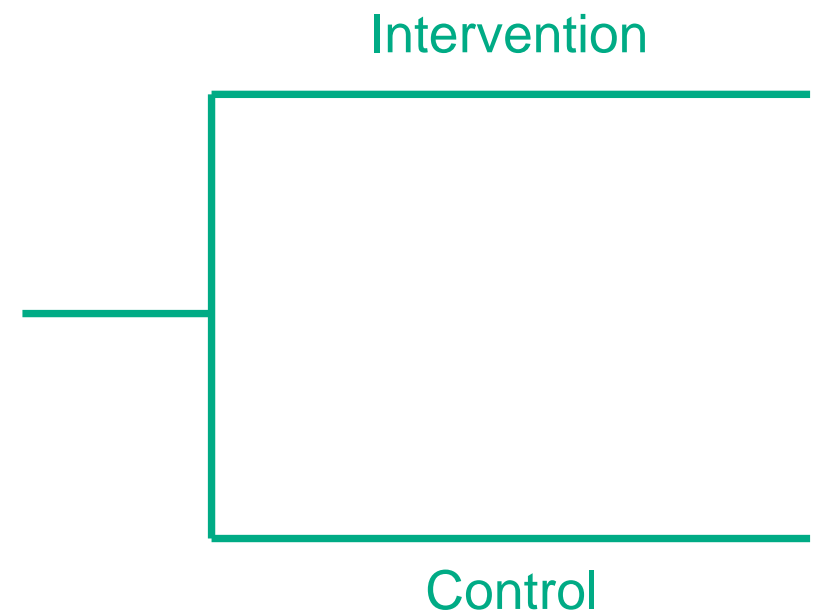
The individual treatment effect (ITE) is defined as a difference in potential outcomes

But only one of them is realised for each person

So, we can't calculate the ITE 🤔

We can estimate an **average treatment effect** with RCT or QED 🎈

That is the average difference in potential outcomes(!)

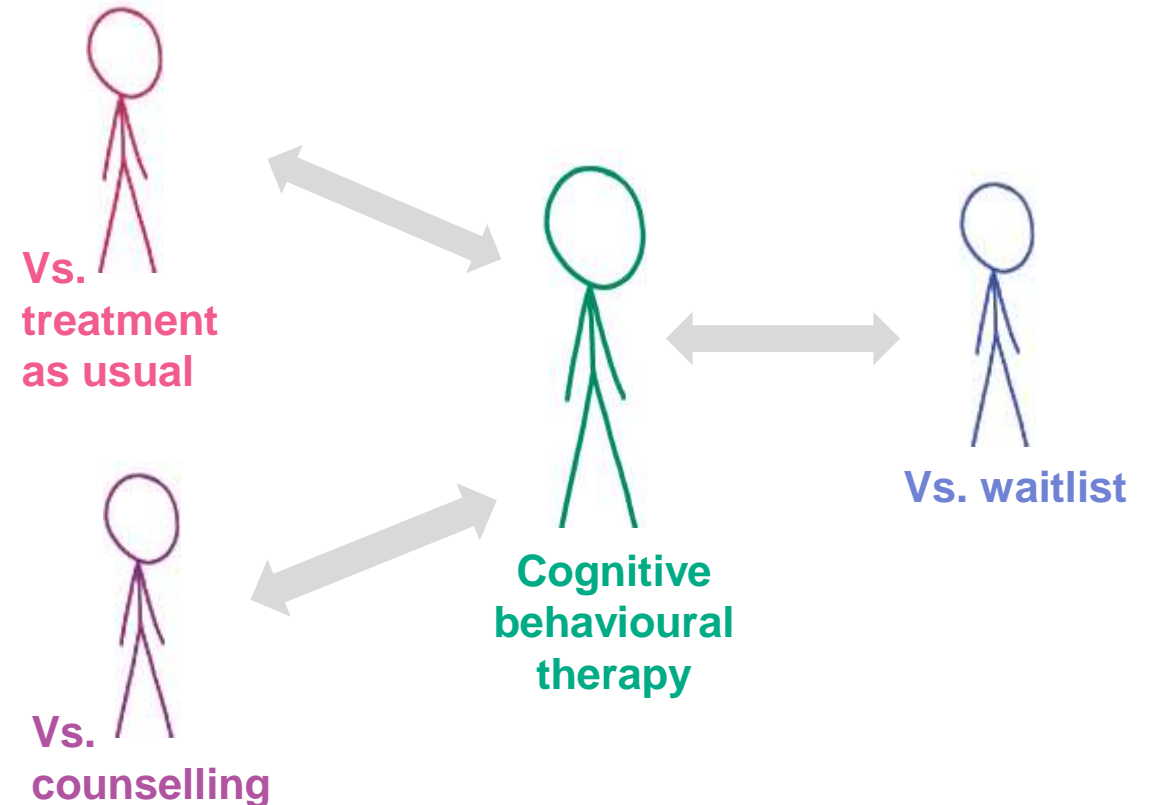


Yet another problem needing theory: what does our average treatment effect estimate represent?

Average treatment effects (ATEs)
are differences

RCTs can give you an unbiased
estimate of the difference

The ATE can't tell you
what made the difference



**Can a counterfactual-
based evaluation be
conducted without a
comparison group?**



Revisiting Contribution Analysis

John Mayne

“In many situations a counterfactual perspective on causality—which is the traditional evaluation perspective—is unlikely to be useful; experimental designs are often neither feasible nor practical...”

“[Contribution analysis] uses a stepwise (generative) **not a counterfactual approach** to causality.”

Mayne (2019, pp. 173–4)

The Qualitative Method of Impact Analysis

LAWRENCE B. MOHR.

“A design whose purpose is to determine impact will be considered **qualitative** if it relies on **something other than evidence for the counterfactual** to make a causal inference.”

Mohr (2019, p. 71)

A Contribution to Current Debates in Impact Evaluation

Howard White

International Initiative on Impact Evaluation (3ie)

“... having no comparison group is not the same as having no counterfactual. There is a very simple counterfactual: what would Y have been in the absence of the intervention? The counterfactual [*in a pre-post example given*] is that it would have remained unchanged, that is, the same as before the intervention.”

White (2010, p. 157)

The Counterfactual Definition of a Program Effect

Charles S. Reichardt¹ 

“I argue that counterfactual thinking need not be limited to a narrow range of methods [...]. [...] Counterfactual thinking can be used just as well as the basis for a wide range of other methods for assessing program effects, including qualitative methods.”

Reichardt (2022, p. 159)

What is a counterfactual? Examples:

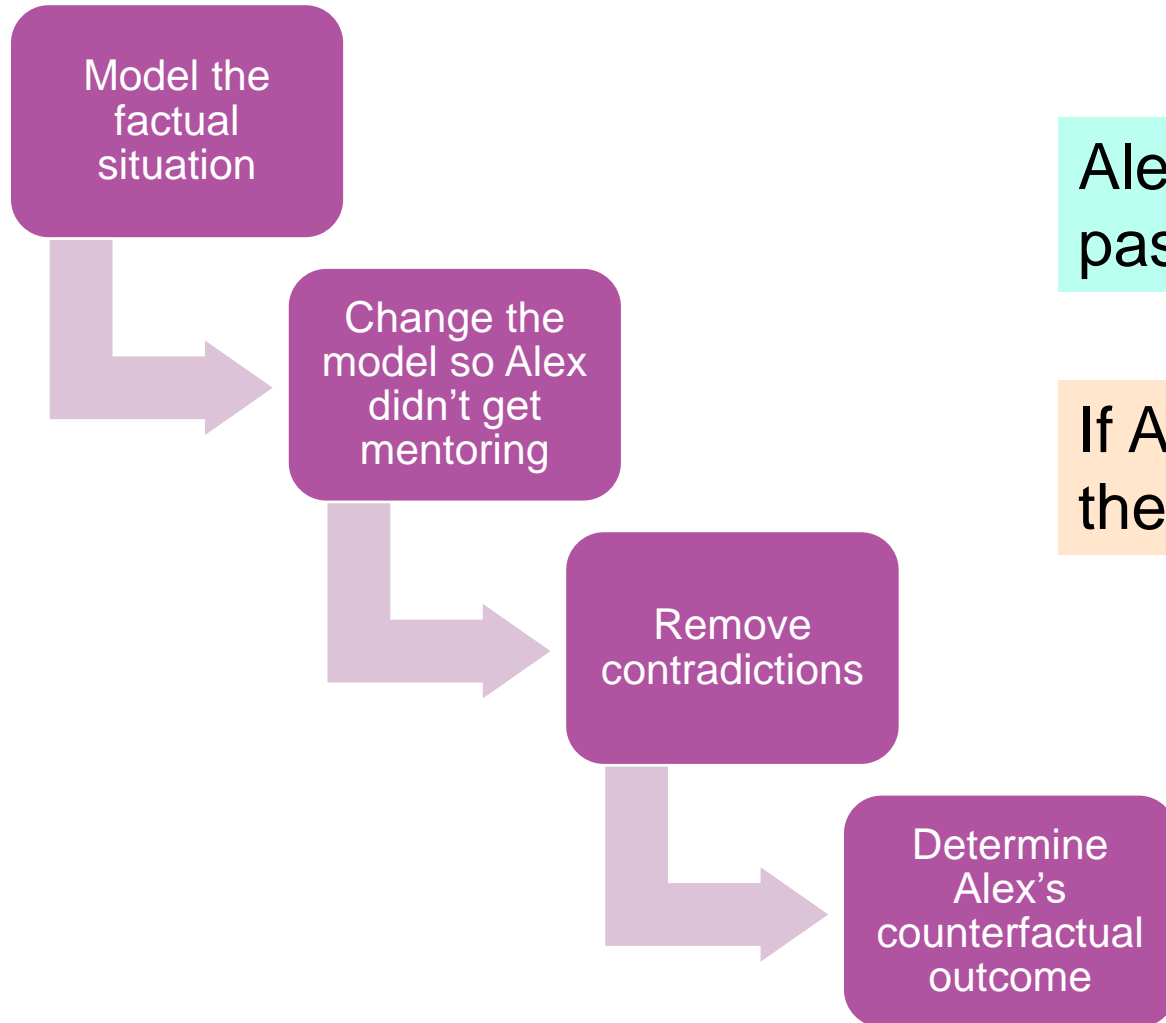
If Oswald hadn't killed Kennedy,
someone else would have

If mentoring programme participants had been
assigned to control, their outcomes would
have been 20 points worse on average

If control participants had been assigned to
mentoring, their outcomes would have
been 20 points better on average

If X had been x , then Y would have been y

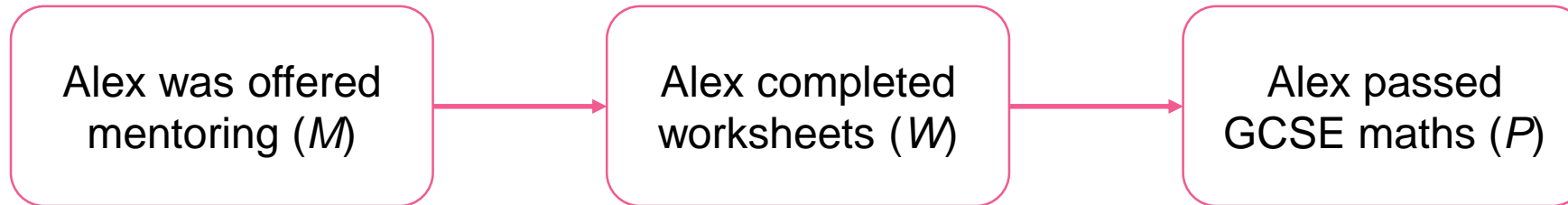
How to reason about counterfactuals (e.g., Stalnaker, 1968, p. 44)

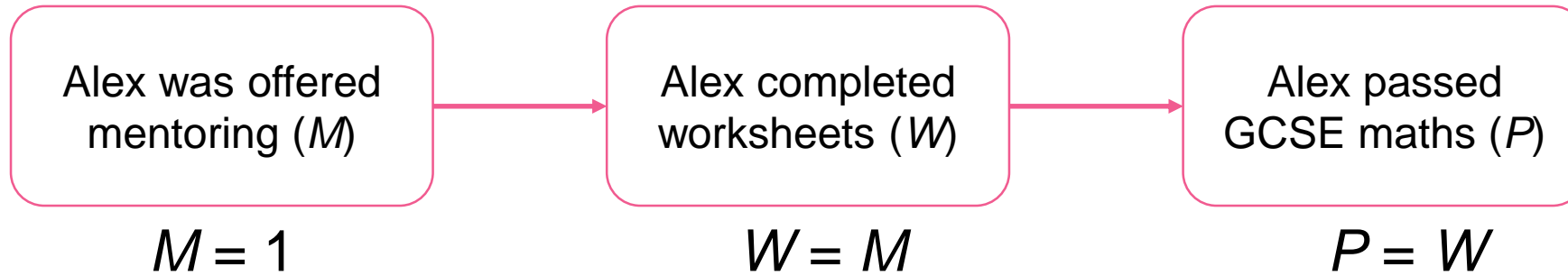


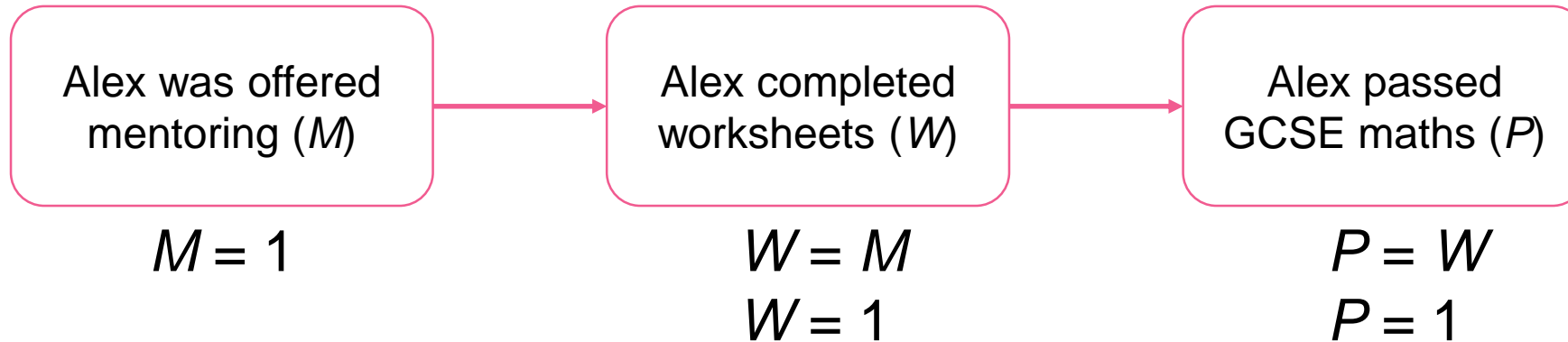
Alex received mentoring and passed GCSE maths

If Alex *hadn't* received mentoring, then they would have failed GCSE maths

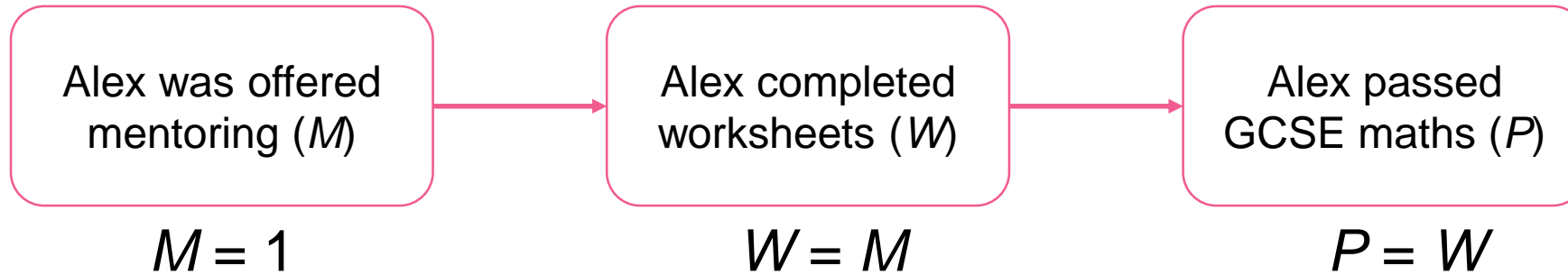
A simple structural equation model



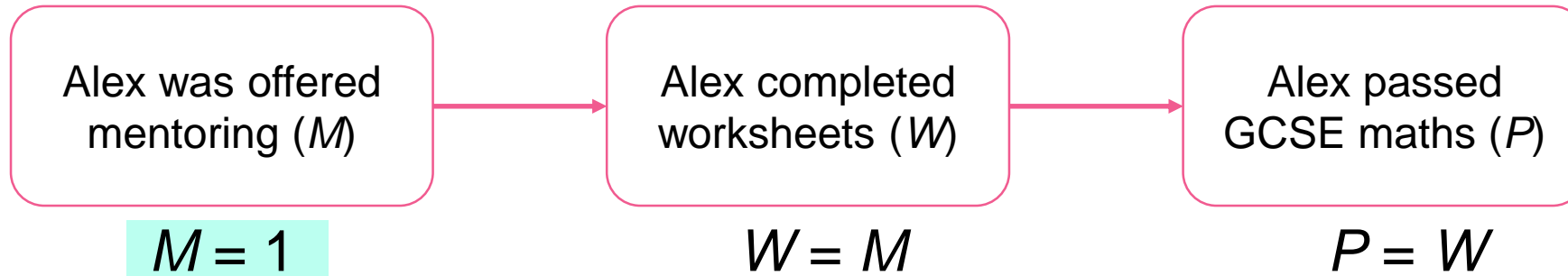




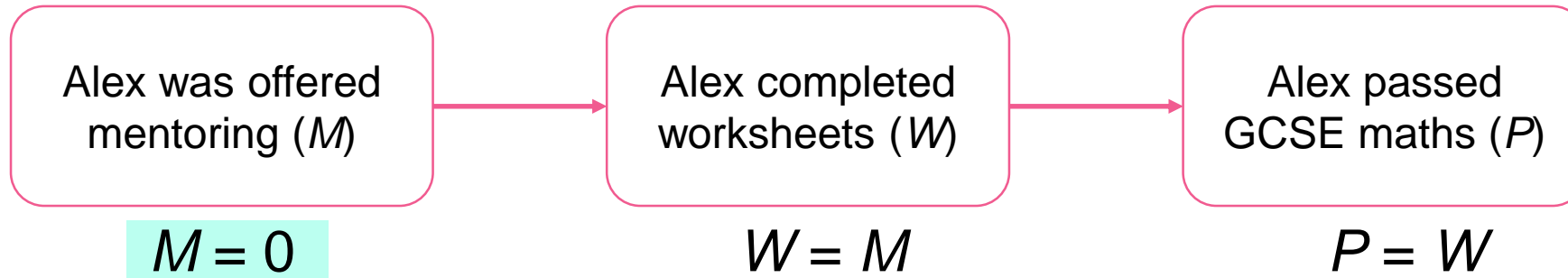
What if Alex hadn't been offered mentoring?



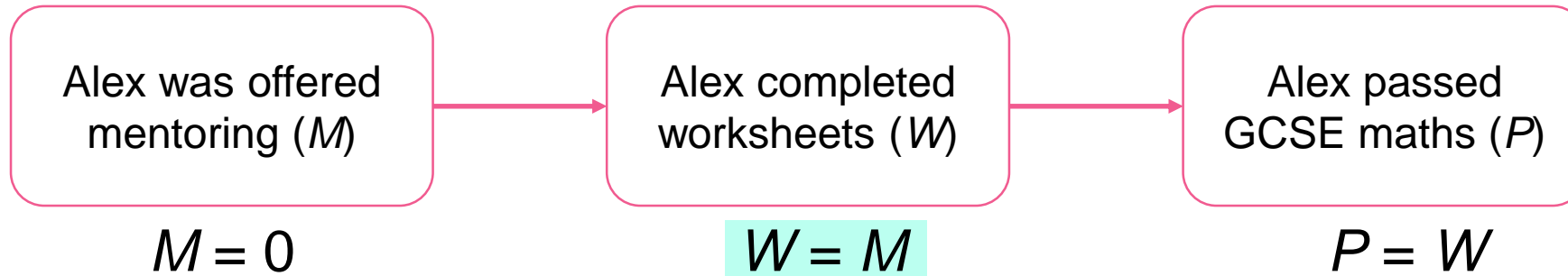
What if Alex hadn't been offered mentoring?



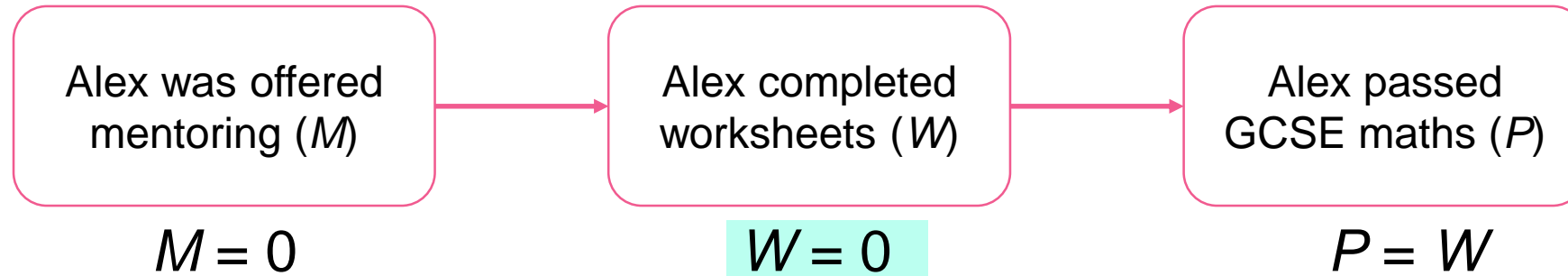
What if Alex hadn't been offered mentoring?



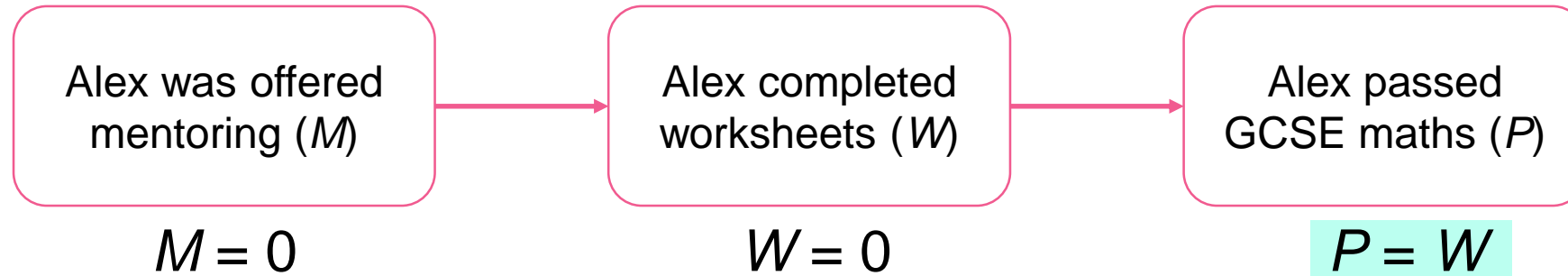
What if Alex hadn't been offered mentoring?



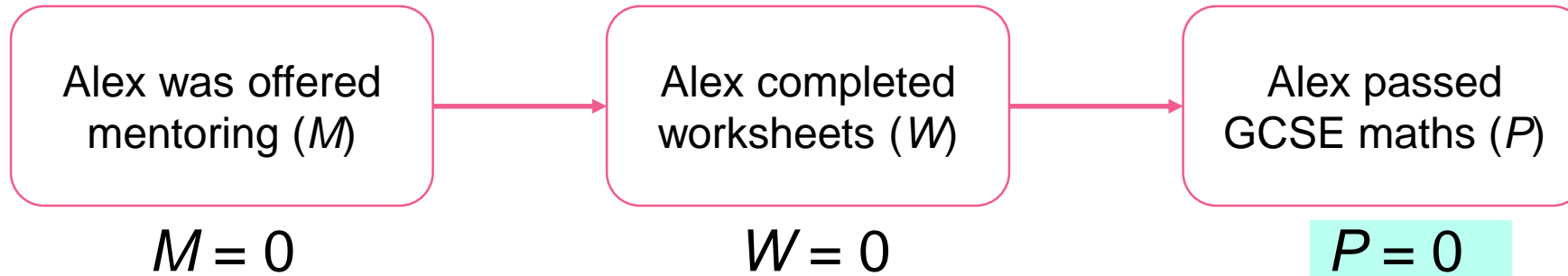
What if Alex hadn't been offered mentoring?

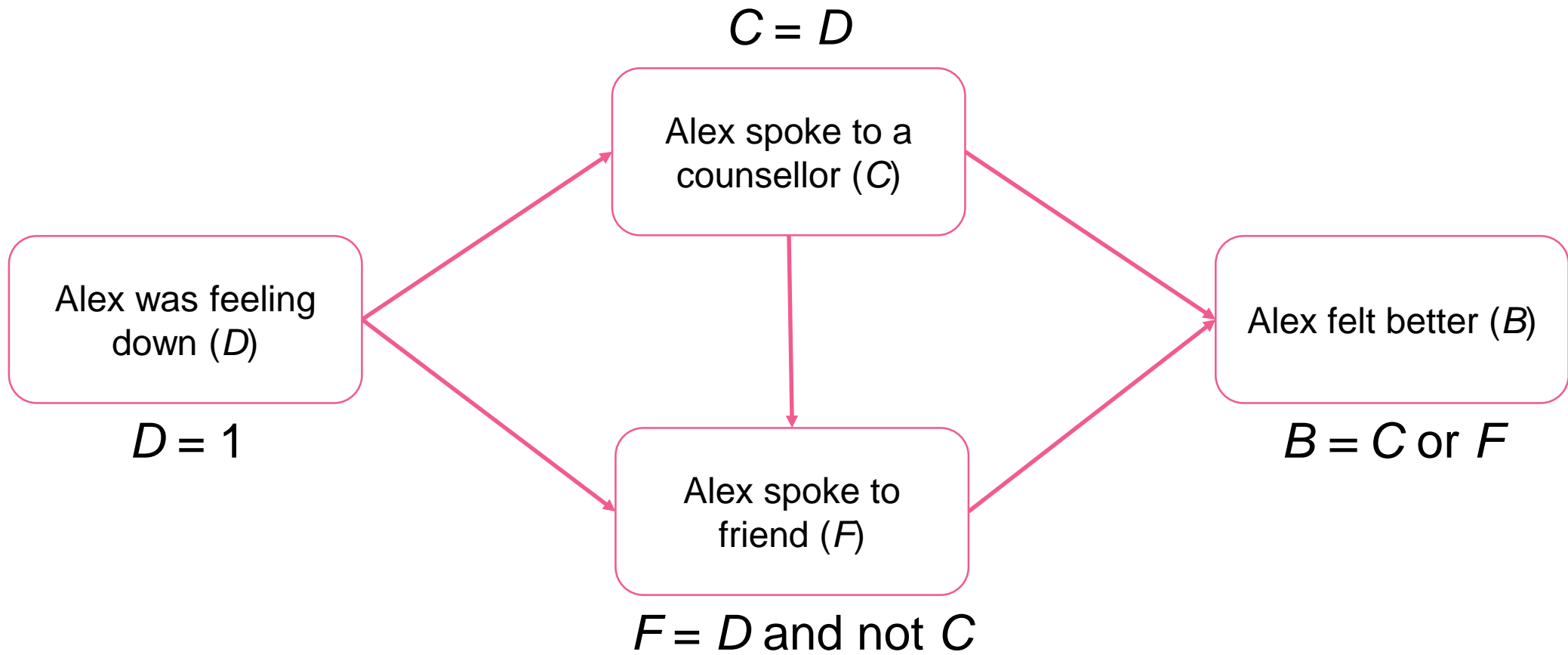


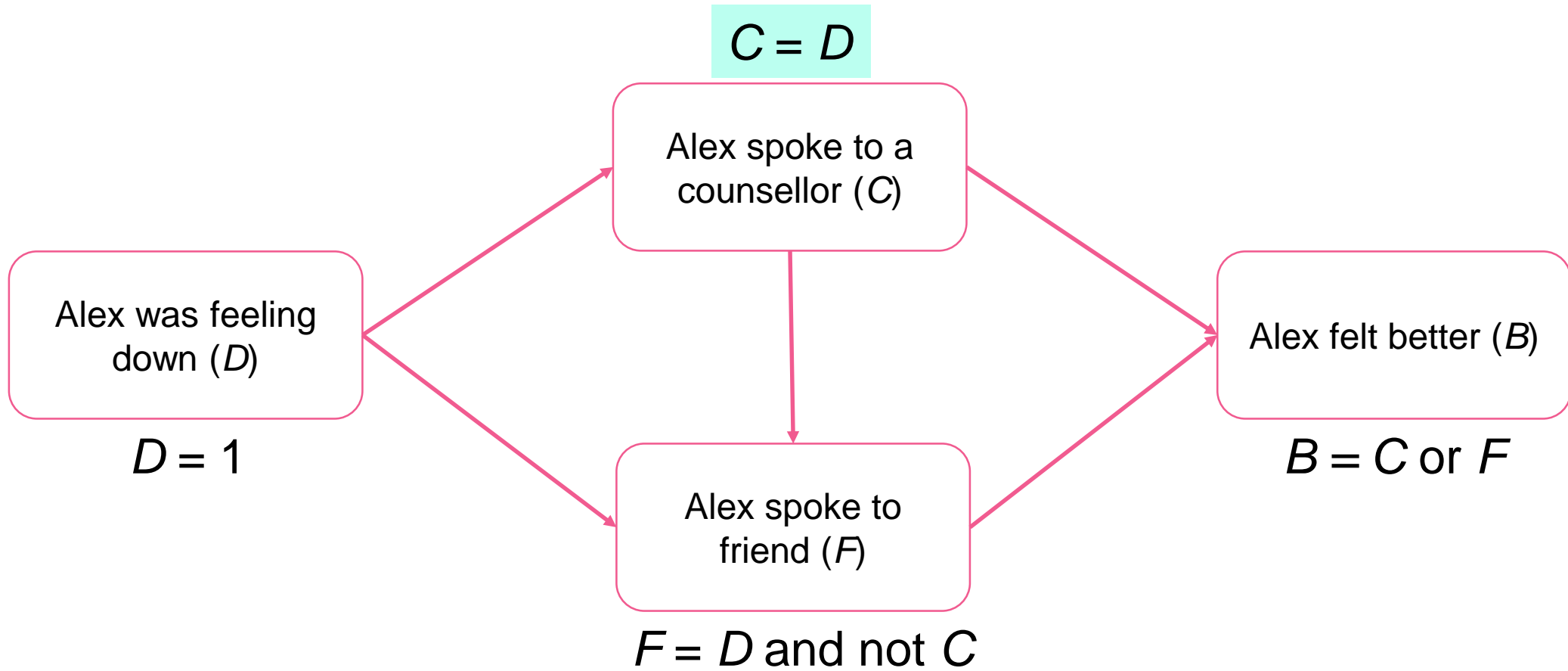
What if Alex hadn't been offered mentoring?

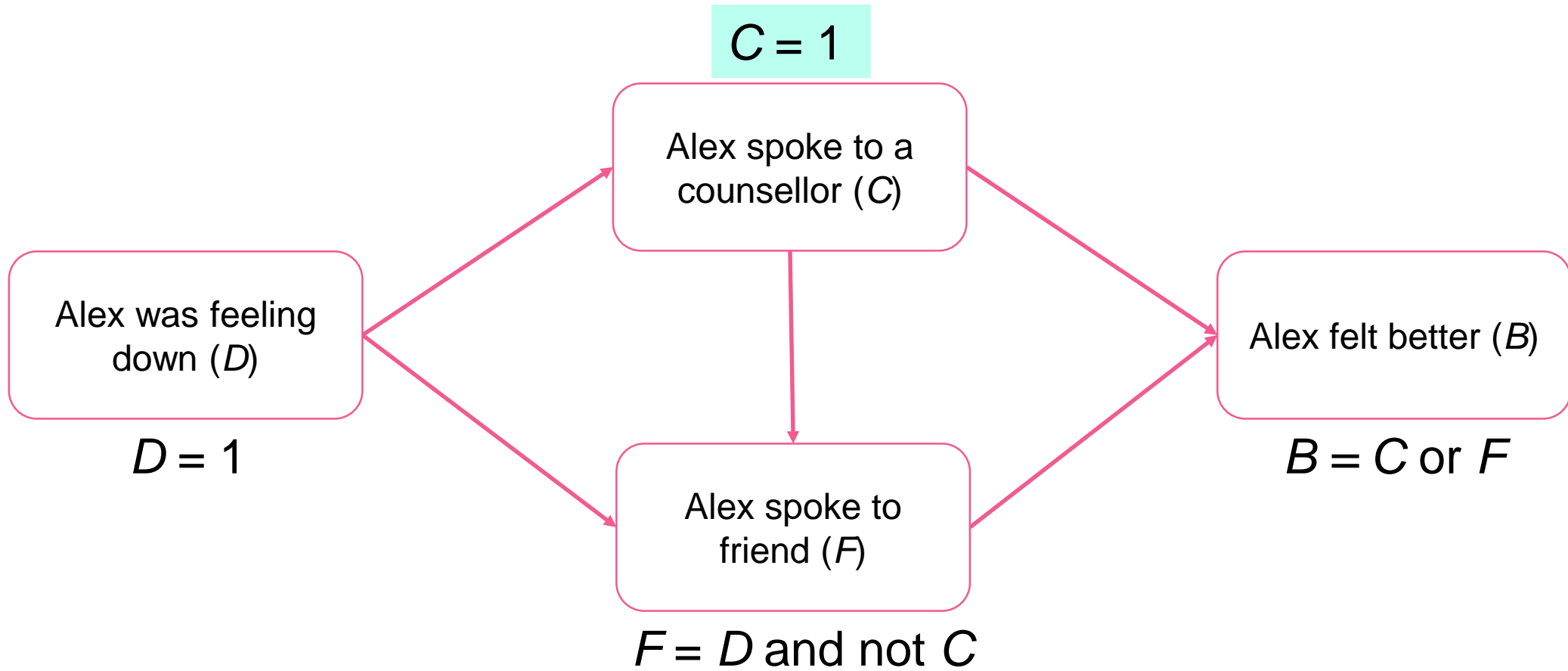


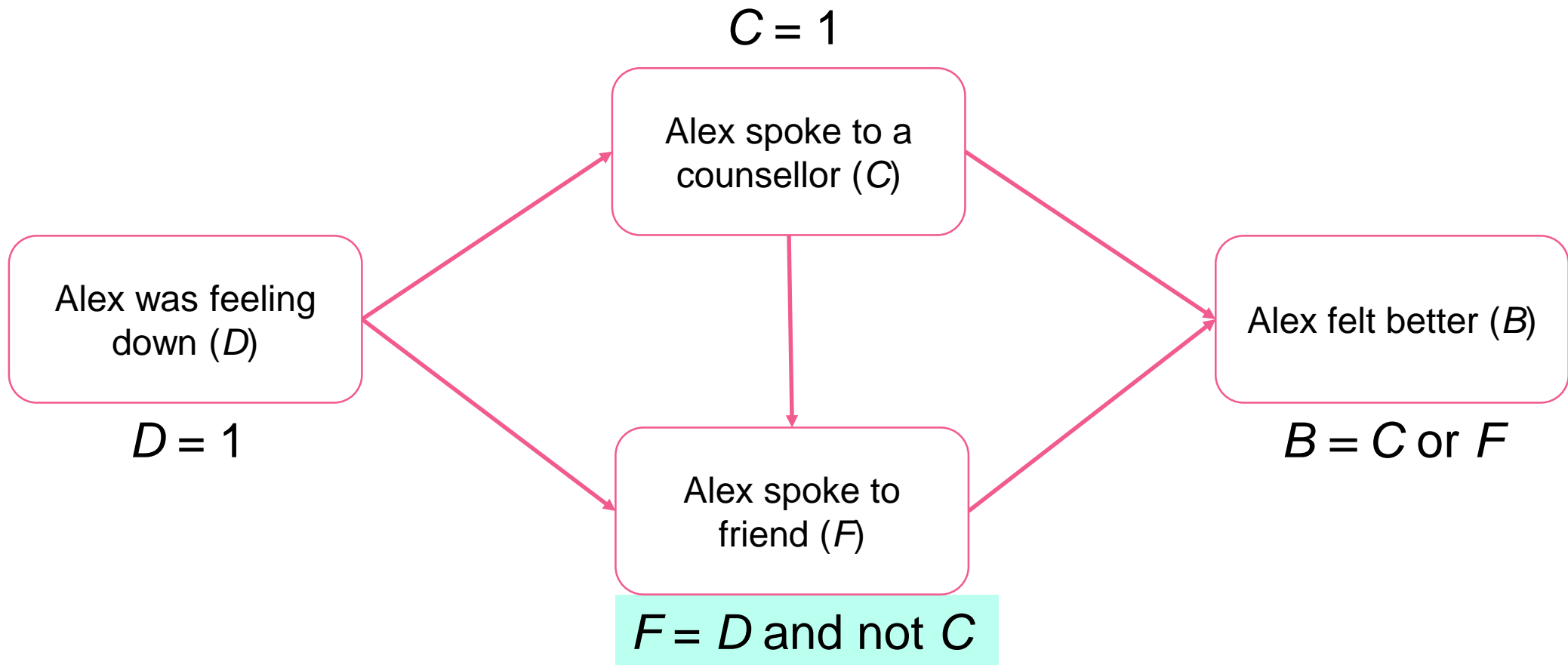
What if Alex hadn't been offered mentoring?

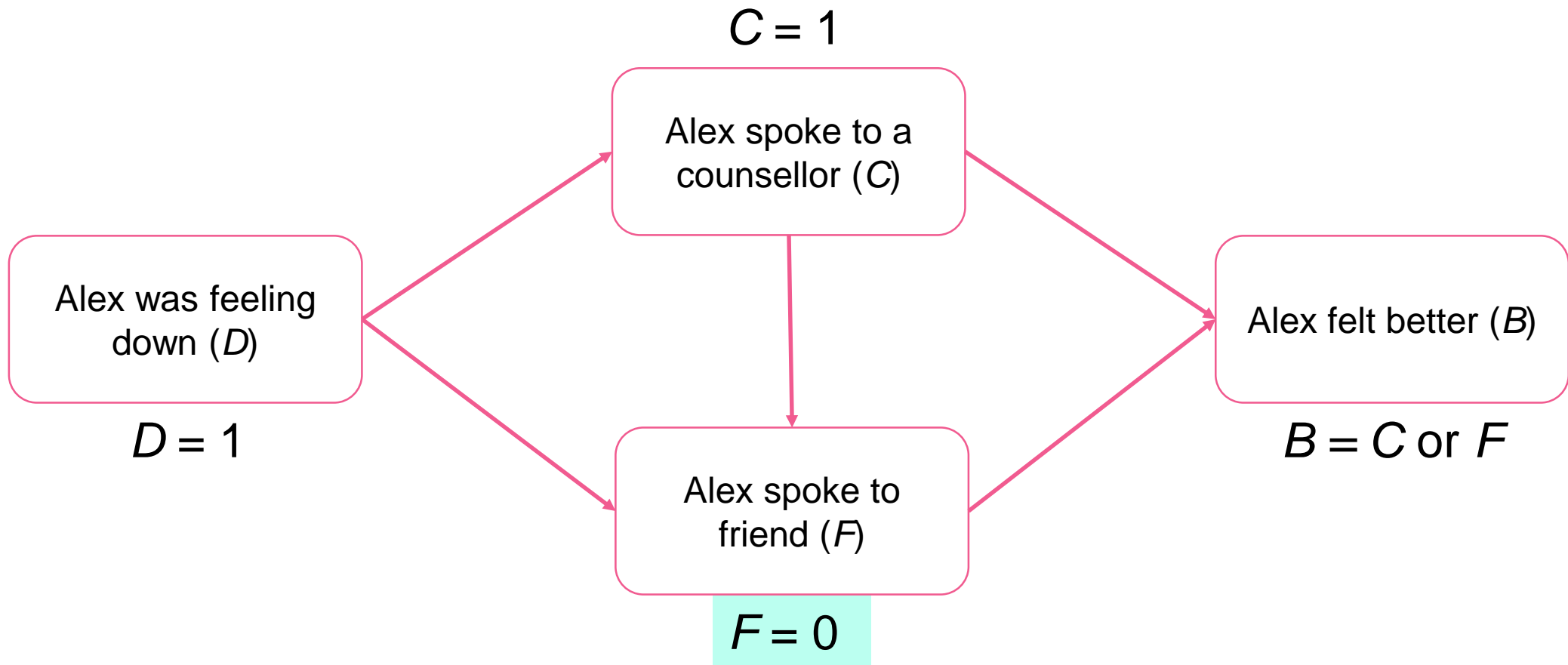


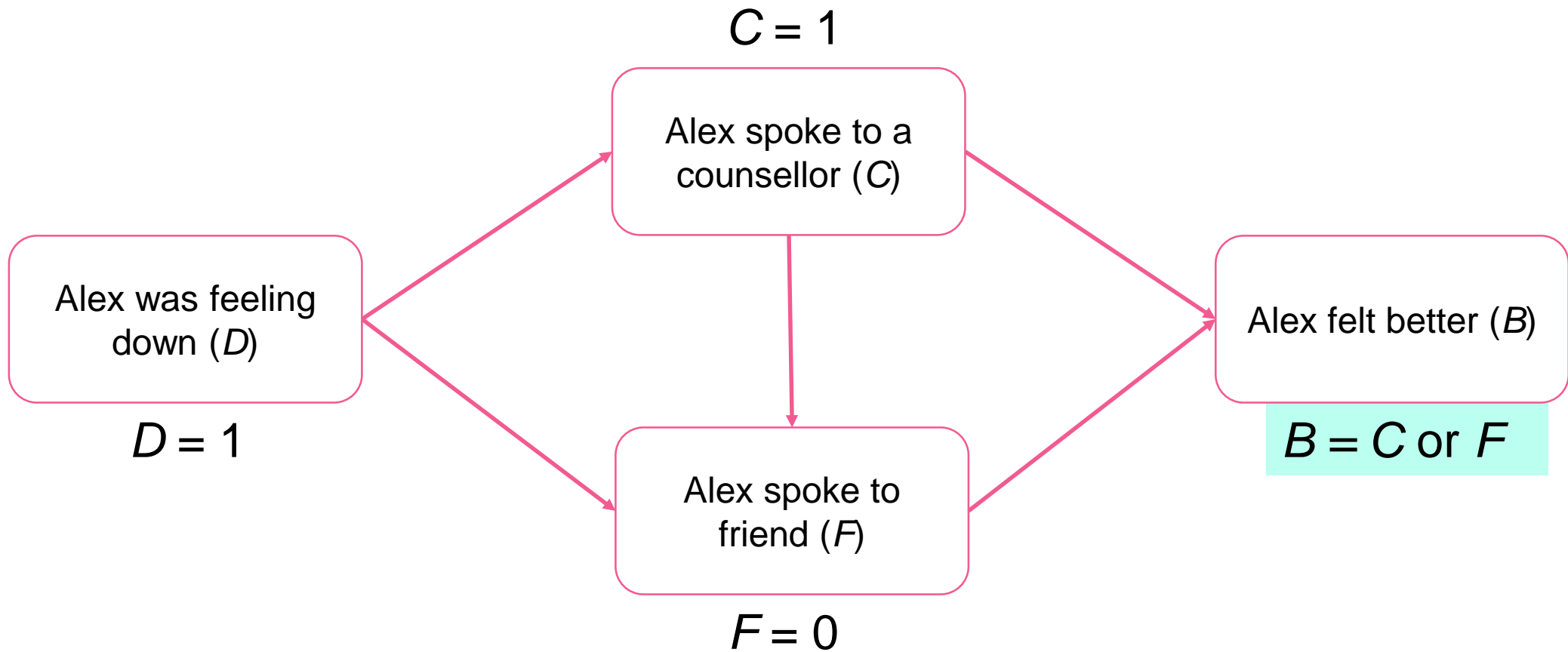


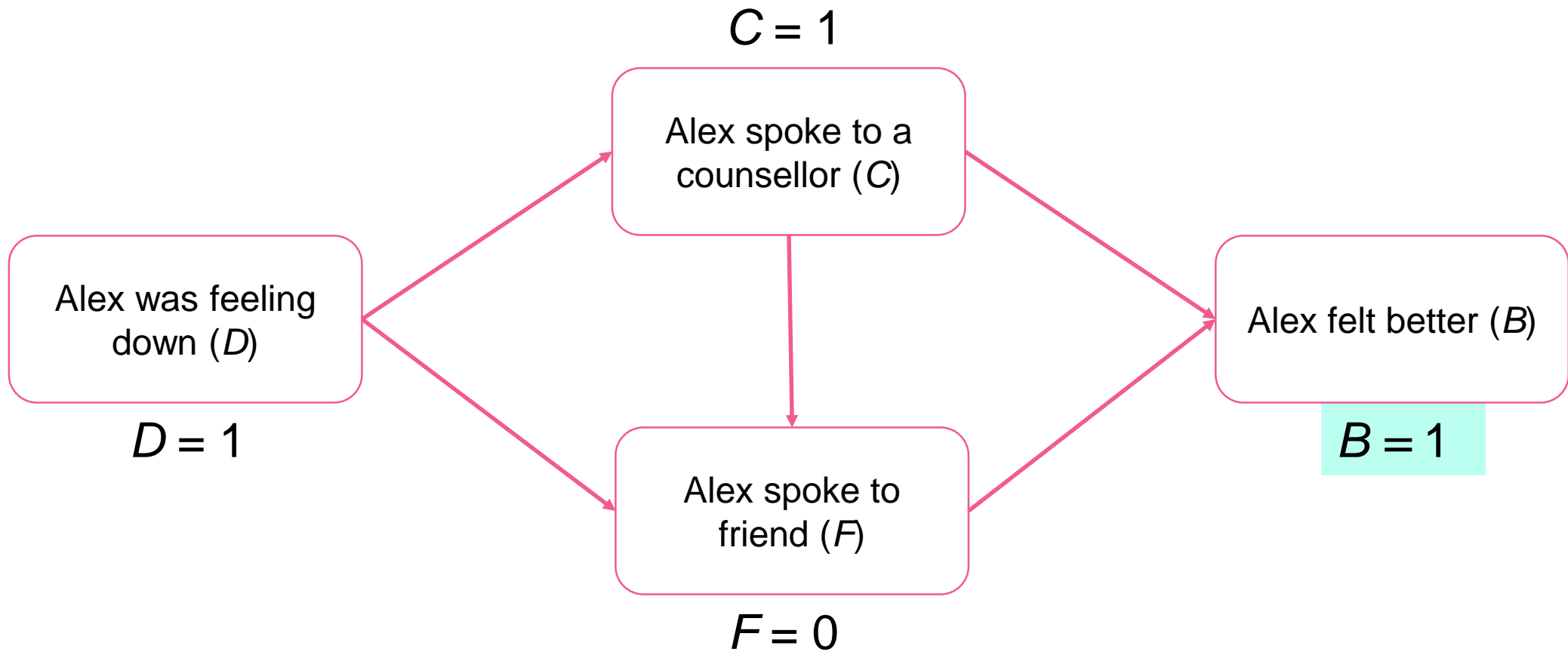


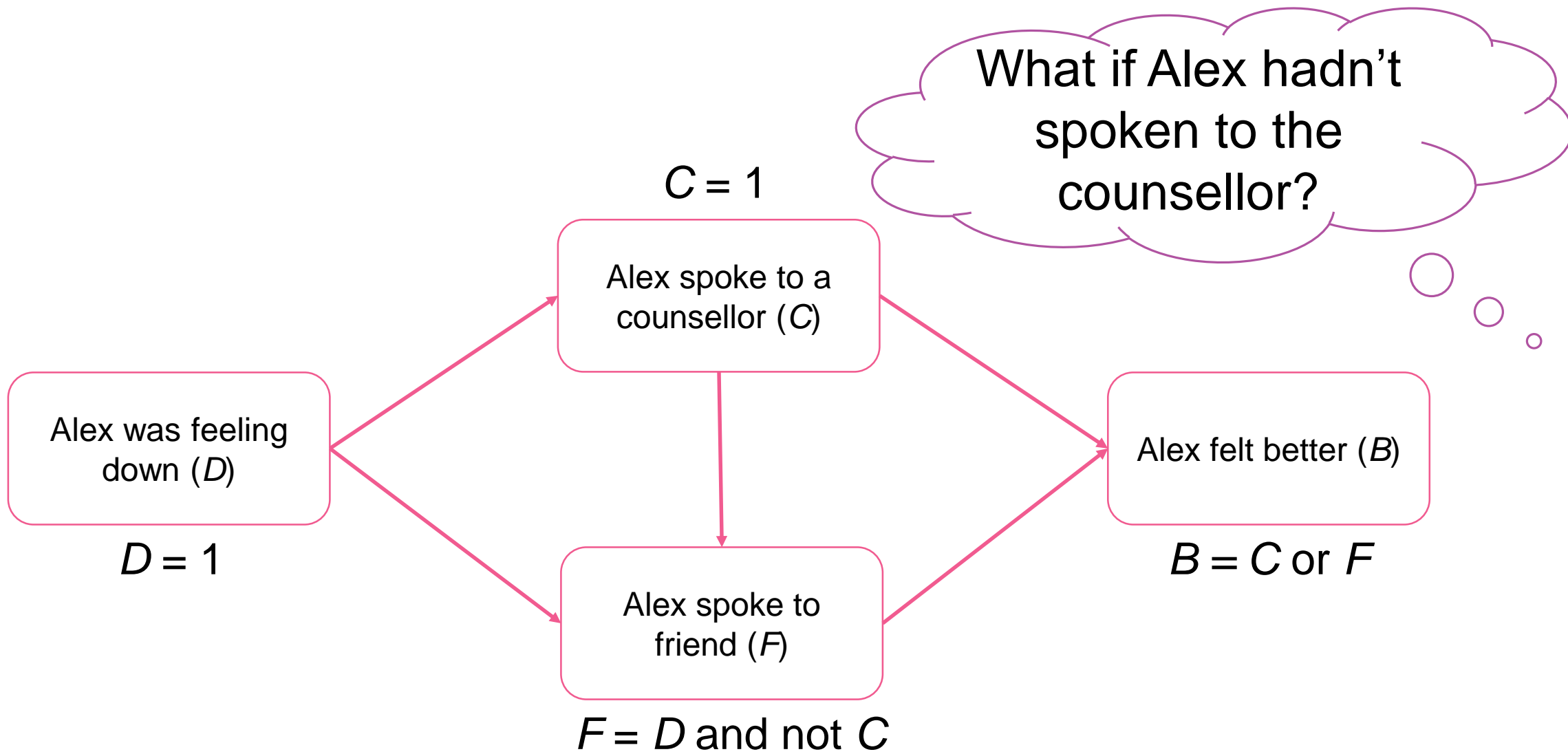


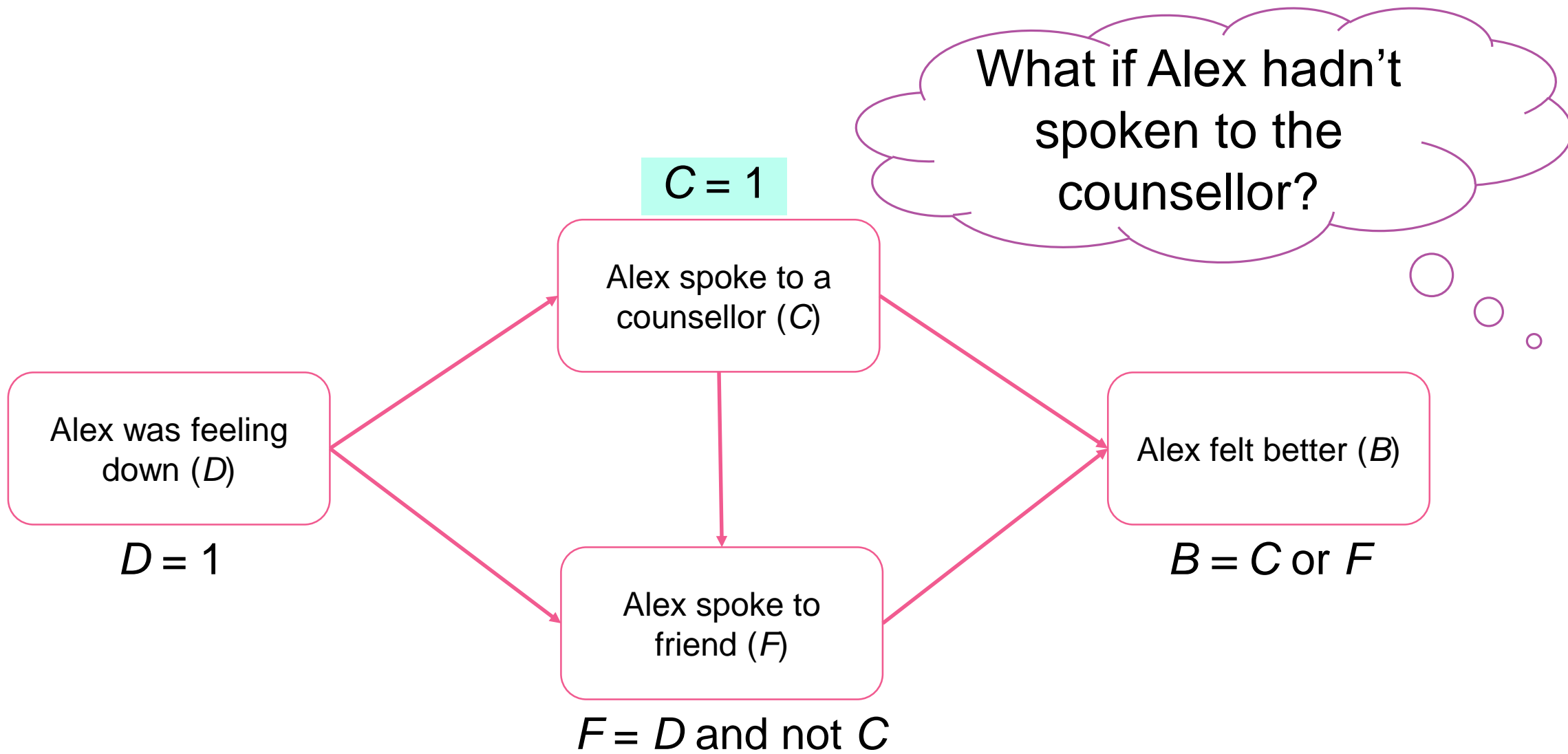


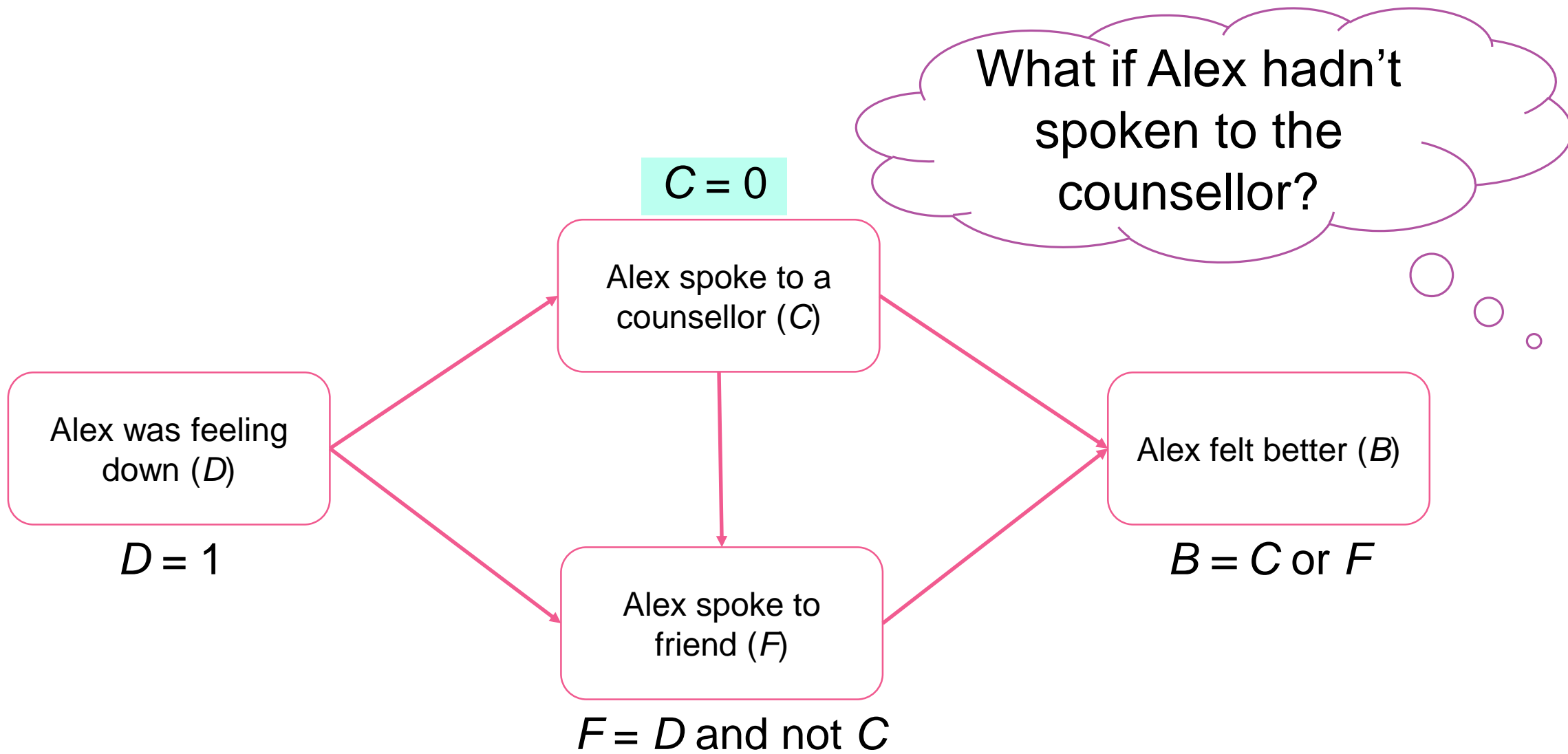


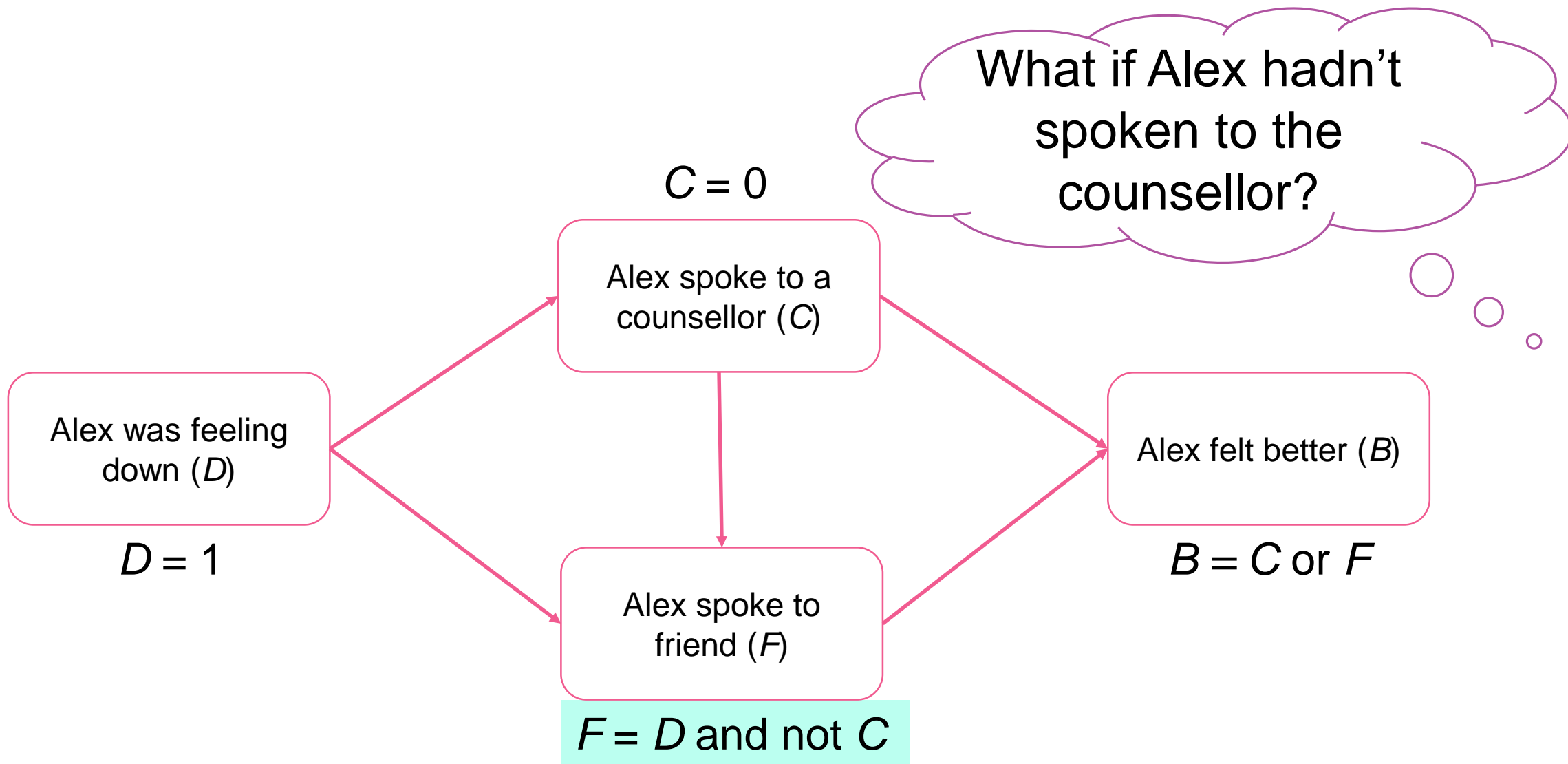


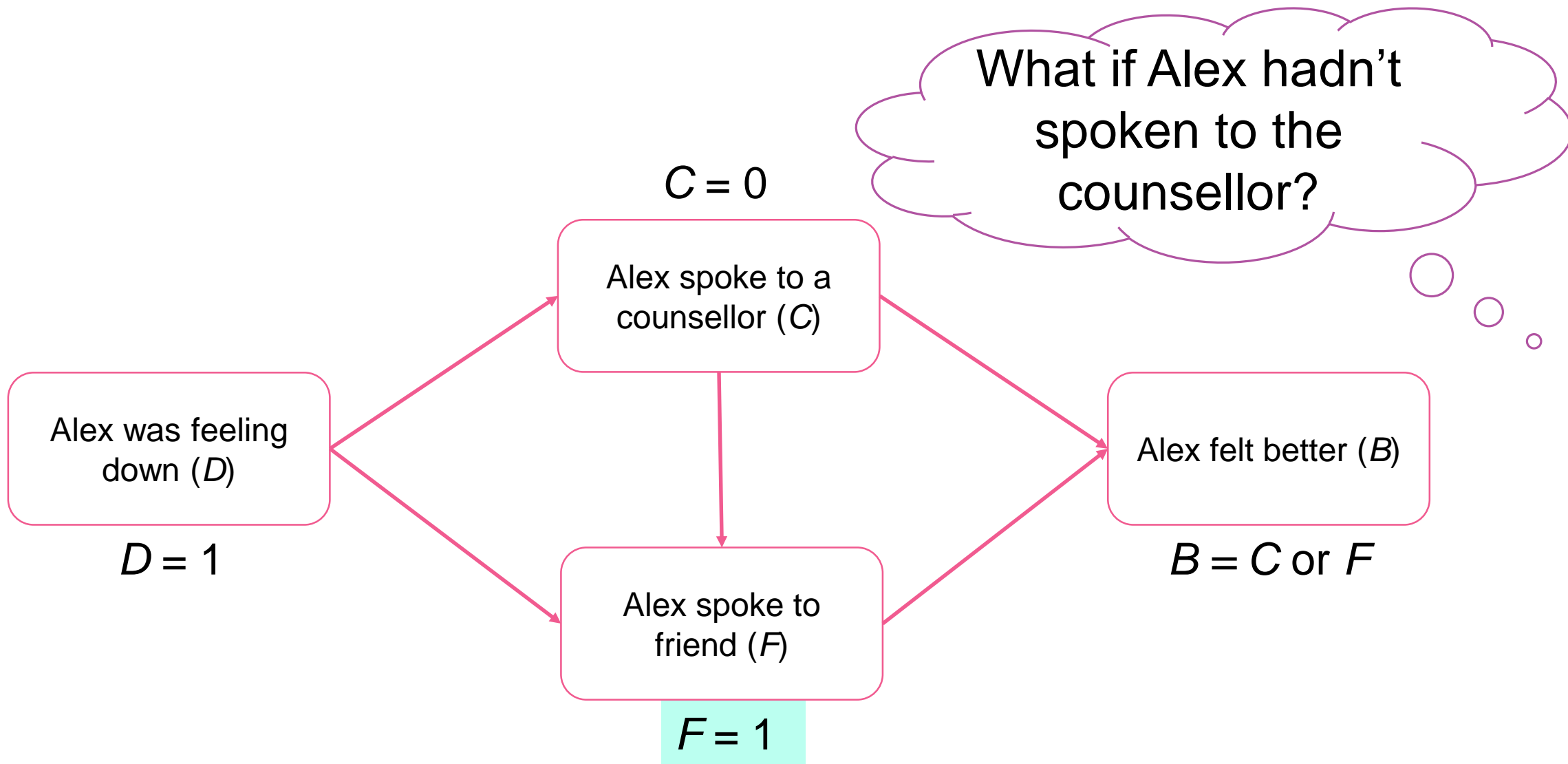


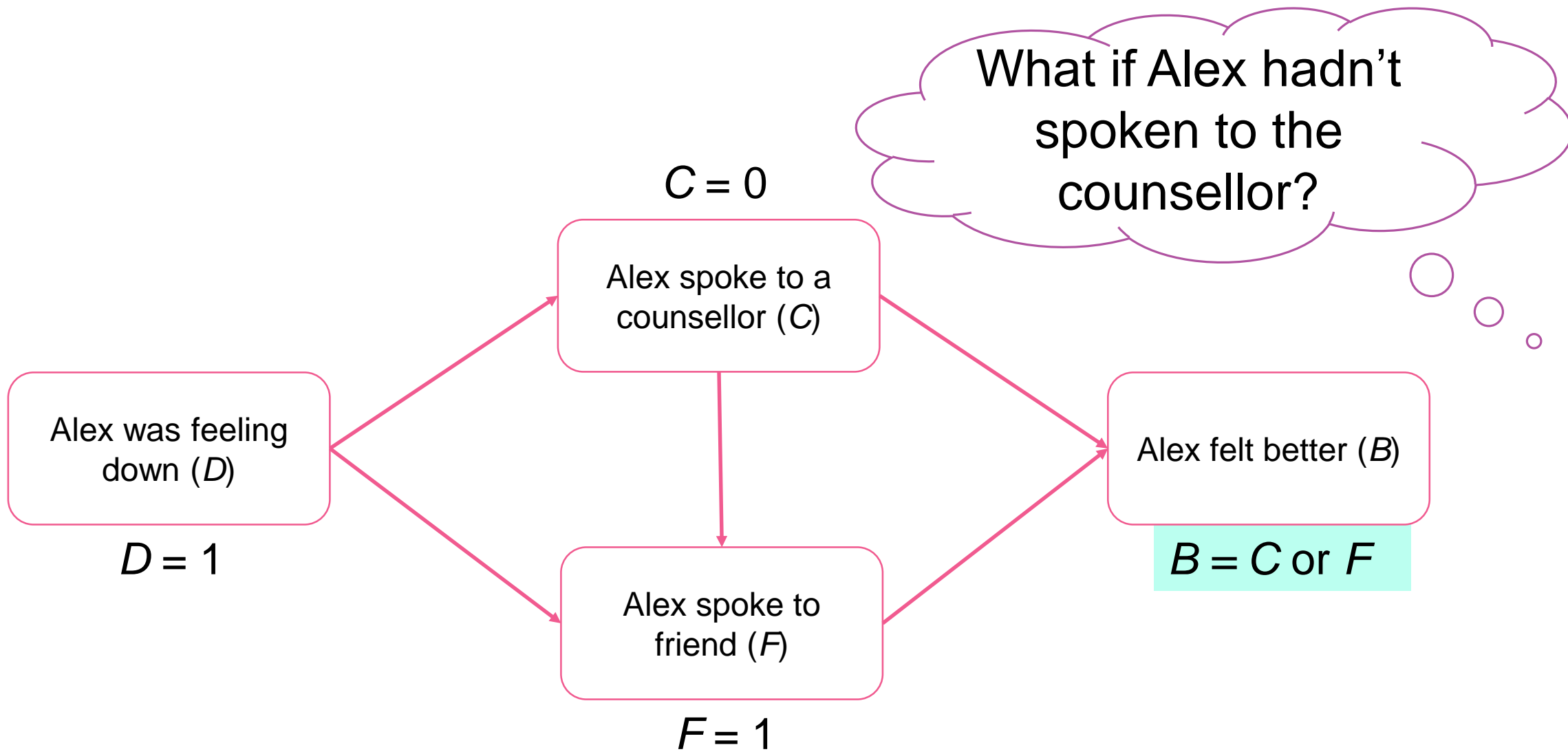


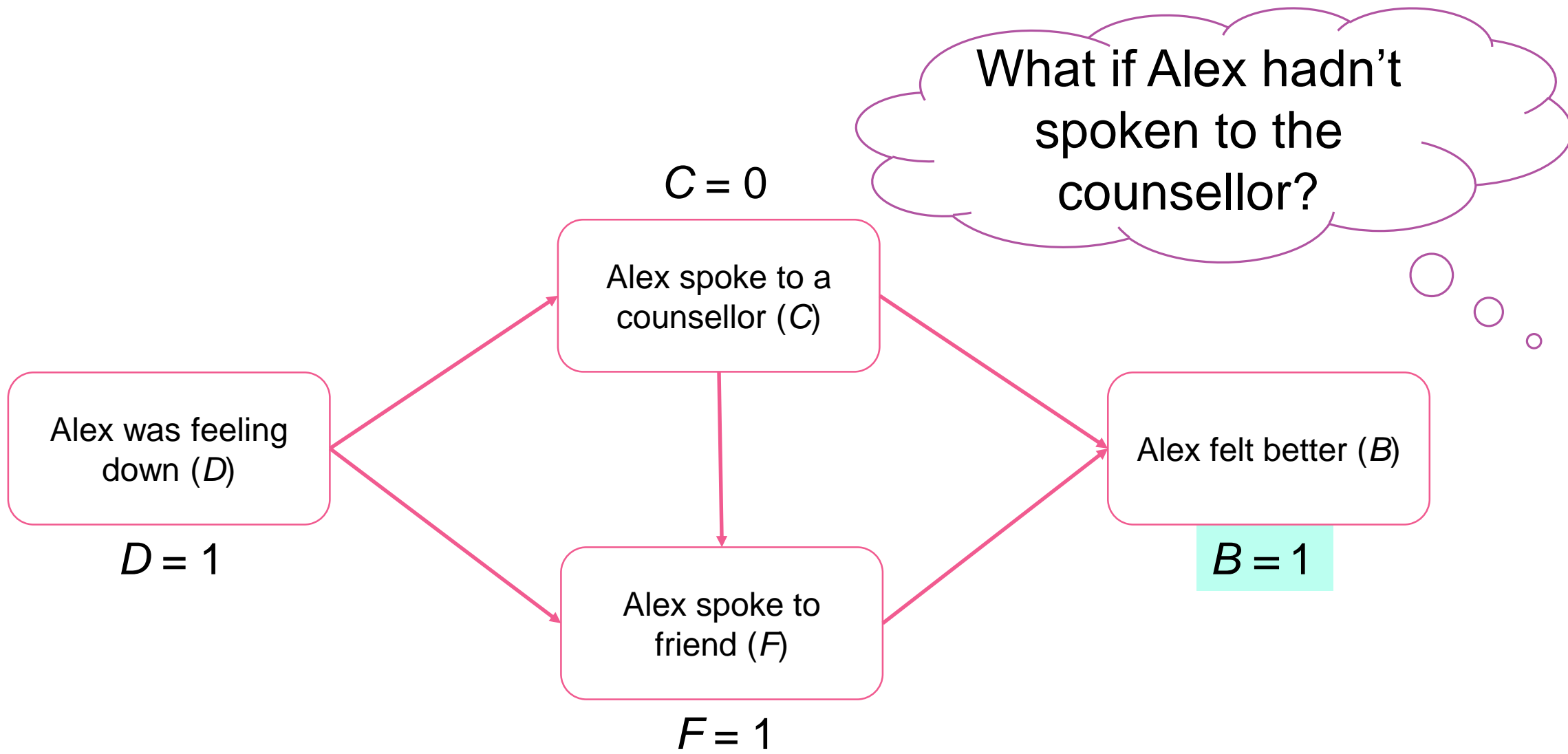












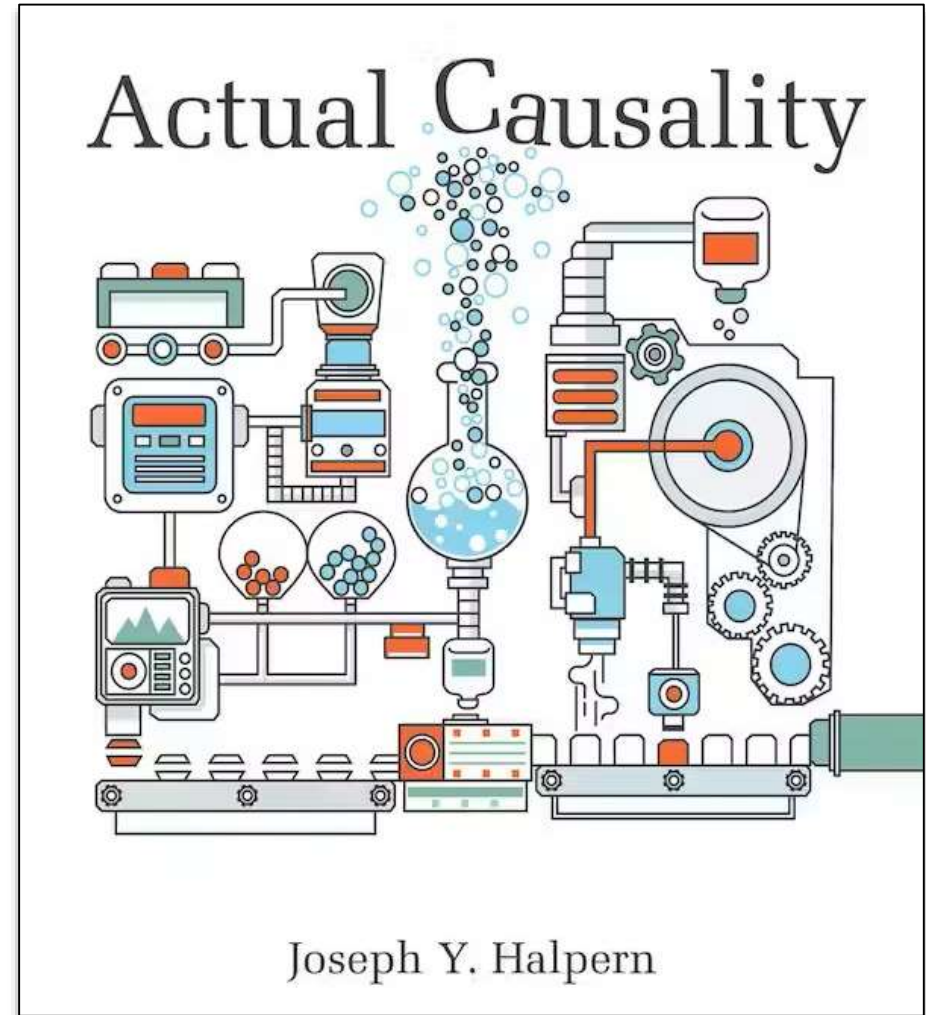
Did going to the counsellor help Alex feel better...?

Factual: When Alex went to the counsellor, they felt better

Counterfactual: When Alex didn't go to the counsellor, they spoke to a friend, and felt better

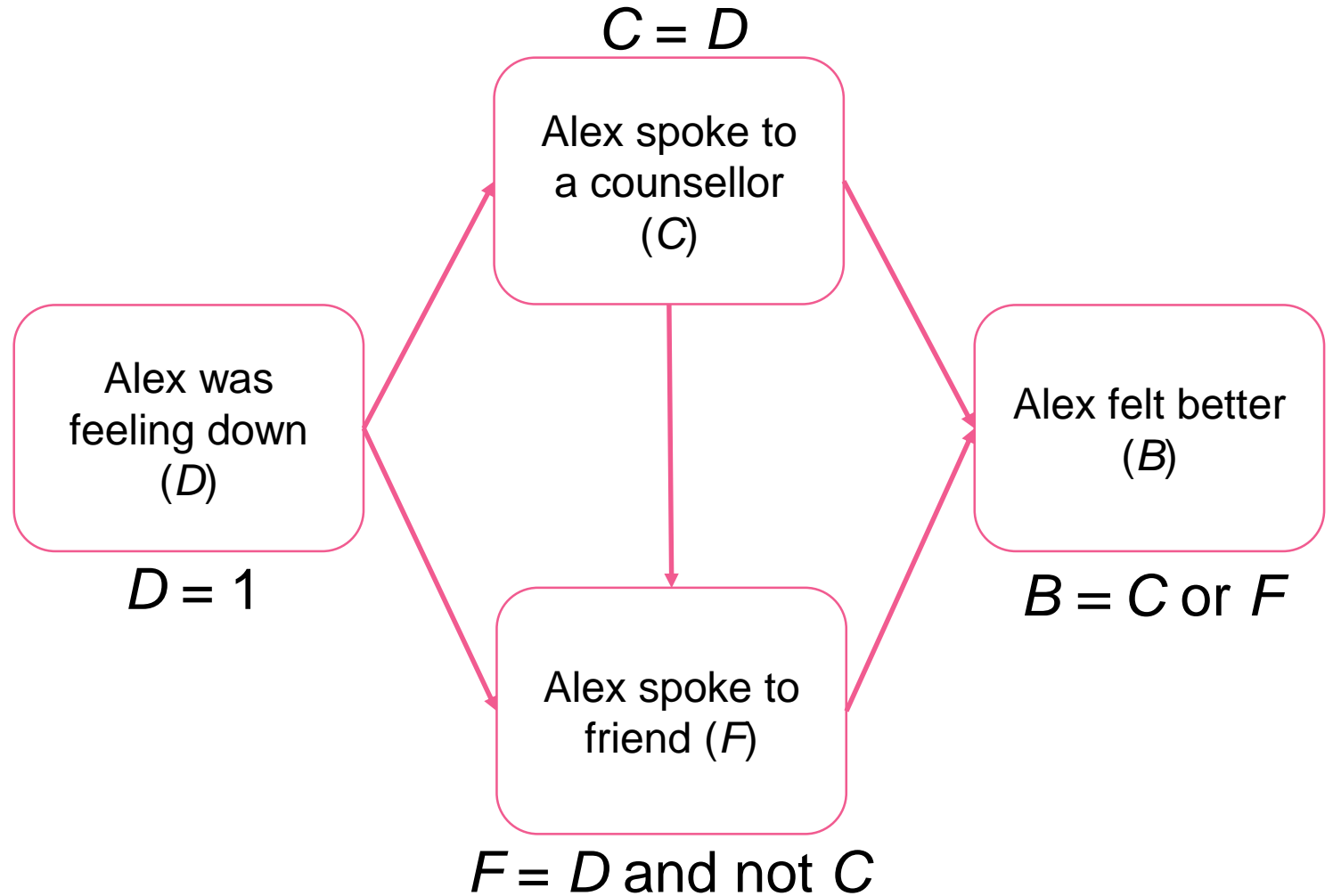
Based on the counterfactual, going to the counsellor didn't make a difference! 🤖

Halpern-Pearl causality helps with examples like these...



Is speaking to a counsellor an *actual cause* of Alex feeling better?

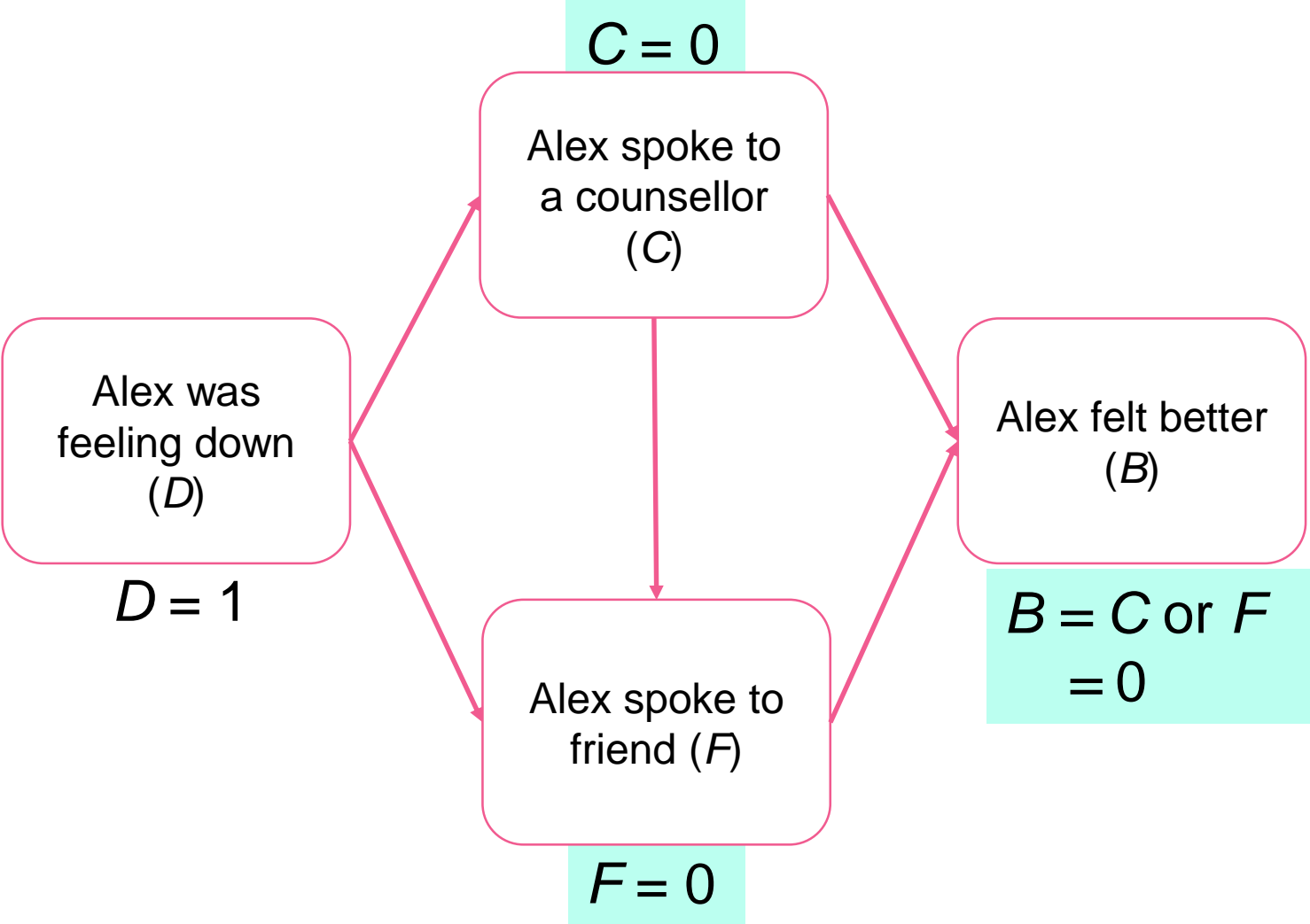
AC2(a). Is there a setting for some subset of variables such that when Alex didn't speak to the counsellor, they didn't feel better?



Is speaking to a counsellor an *actual cause* of Alex feeling better?

AC2(a). Is there a setting for some subset of variables such that when Alex didn't speak to the counsellor, they didn't feel better?

✓ $F = 0$

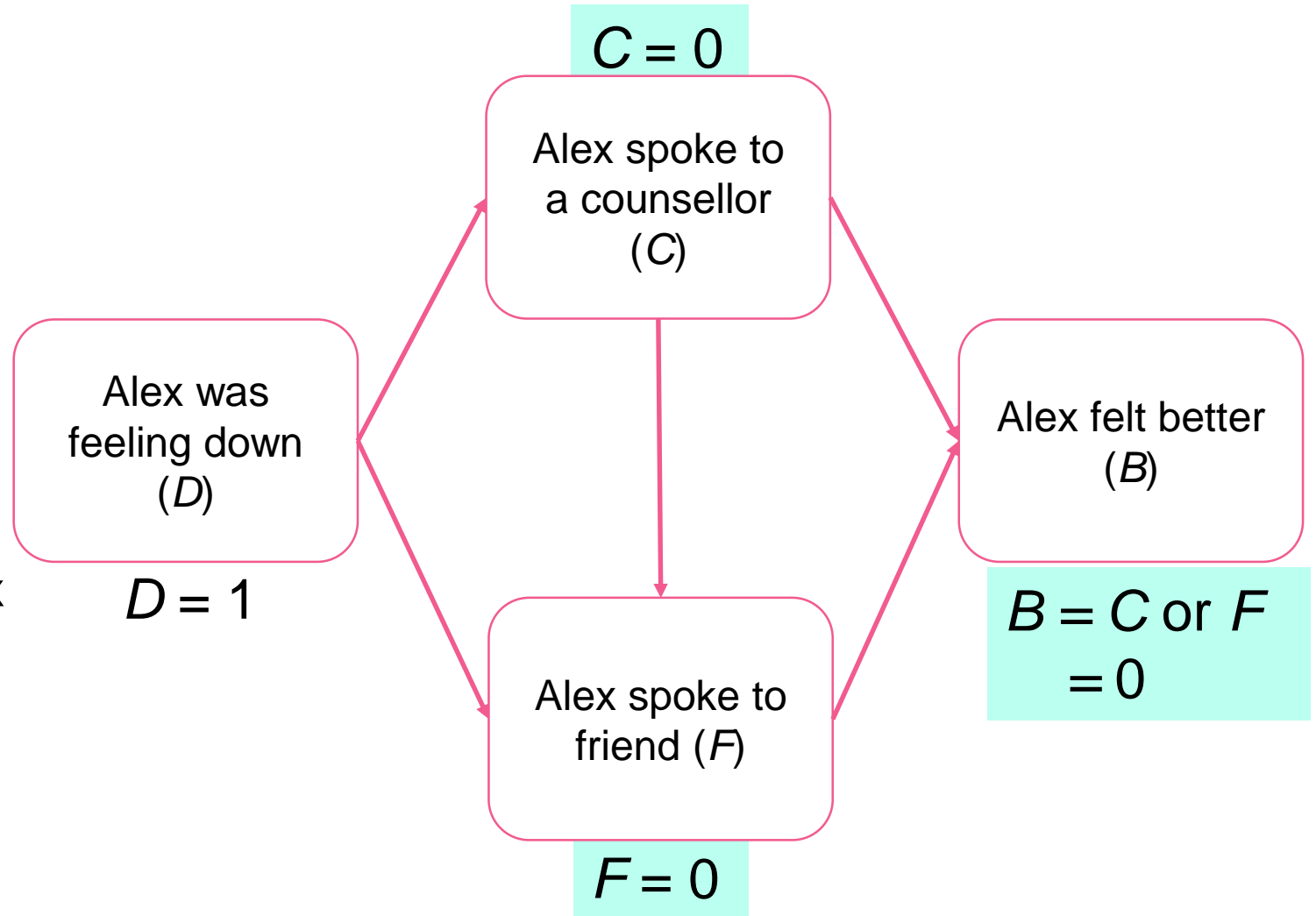


Is speaking to a counsellor an *actual cause* of Alex feeling better?

AC2(a). Is there a setting for some subset of variables such that when Alex didn't speak to the counsellor, they didn't feel better?

✓ $F = 0$

AC2(b). Holding $F = 0$, does speaking to the counsellor help Alex feel better again?



Is speaking to a counsellor an *actual cause* of Alex feeling better?

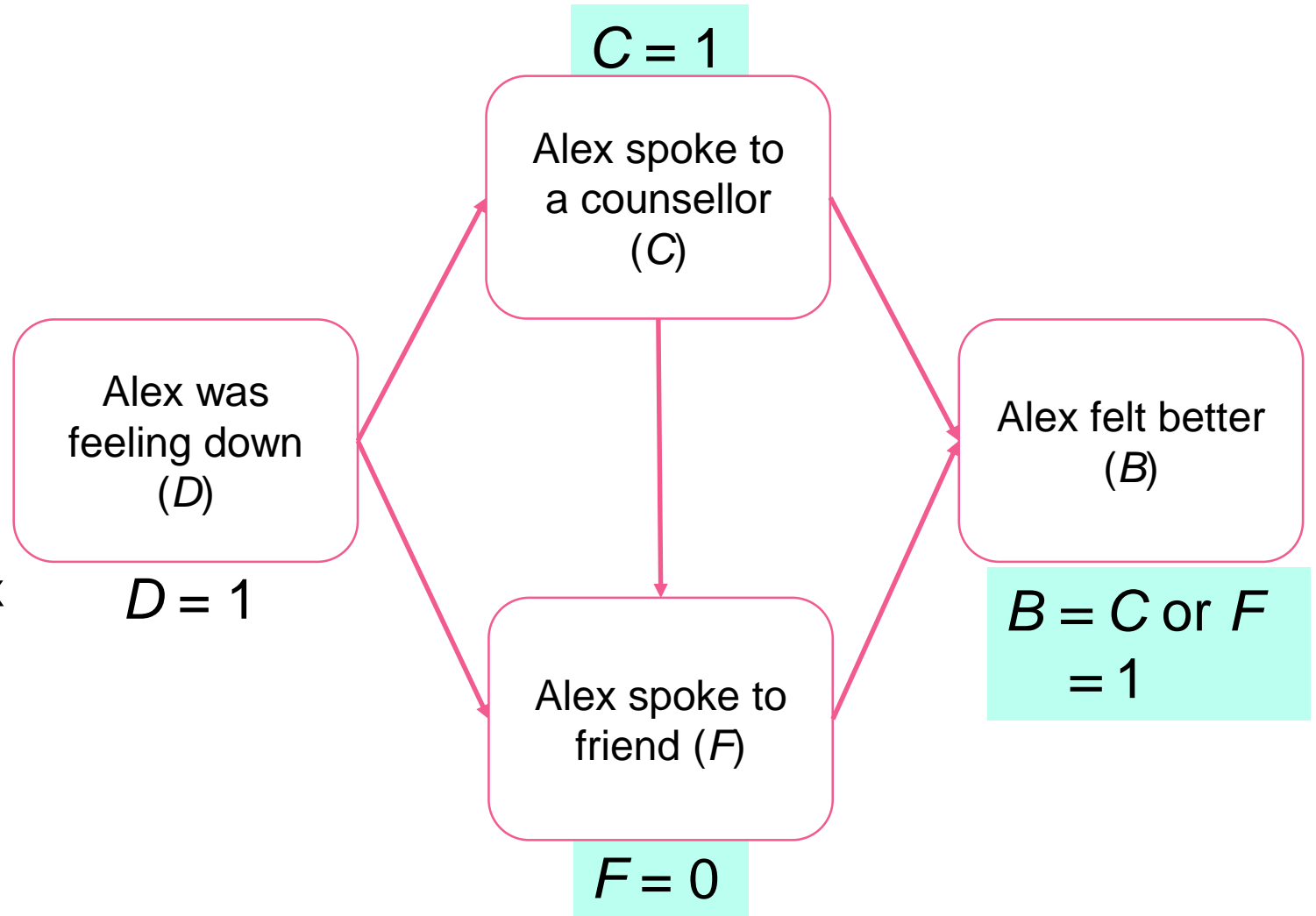
AC2(a). Is there a setting for some subset of variables such that when Alex didn't speak to the counsellor, they didn't feel better?

✓ $F = 0$

AC2(b). Holding $F = 0$, does speaking to the counsellor help Alex feel better again?

✓

(This is the gist of how AC2 works)



RCTs and quasi-experiments are highly theory-dependent

Originally included in the genre of theory-based evaluation

Counterfactual queries don't need a comparison group

They do need a model/theory of how facts came about that can be modified

Is any of the foregoing true? Should evaluation practice change...?