Introduction 000	Negative argument	Positive Argument	Recapitulation 0000	References

The Unity of Robustness

Corey Dethier

University of Notre Dame Philosophy Department corey.dethier@gmail.com

> April 19 Pacific APA

> > ◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Introduction	Negative argument	Positive Argument	Recapitulation	References

Introduction

Introduction •00	Negative argument 00000	Positive Argument	Recapitulation 0000	References
Backgroun	d: disunity			

Disunity: there's a difference between robustness (or agreement) in modeling and experimental contexts

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

I think disunity is false; or, that *unity* is true.

Introduction ○●○	Negative argument	Positive Argument	Recapitulation	References
Background	d: disanalogy			

Primary argument for disunity goes back to Cartwright, who points out that:

unlike what is the case when experiments or measurements agree, different models "do not constitute independent instruments doing different things, but rather different ways of doing the same thing: instead of being unrelated, they are often alternatives to one another, sometimes even contradictory" (Cartwright 1991, 153).

Introduction 000	Negative argument	Positive Argument	Recapitulation	References
The talk				

Three arguments for unity:

- Negative argument: mutual inconsistency among model assumptions does not imply that models are alternatives (in the requisite sense).
- 2 Positive argument: given some plausible assumptions, there's a formal equivalence.

3 - Recapitulation: the original intuition can be recaptured.

Introduction	Negative argument	Positive Argument	Recapitulation	References

Negative argument

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

Introduction 000	Negative argument ●0000	Positive Argument	Recapitulation 0000	References
The oppo	sed intuition			

When there are multiple measurements, the tools used aren't in conflict (in any sense).

The same isn't true of the models employed in an ensemble of models: these usually rely on assumptions that mutually inconsistent.

▲ロト ▲帰 ト ▲ ヨ ト ▲ ヨ ト ・ ヨ ・ の Q ()

Introduction 000	Negative argument 0●000	Positive Argument	Recapitulation 0000	References
What's v	vrong with this	argument?		

Simply: assumes that what matters for our confidence in model results is whether the model has true assumptions, when what really matters is whether it can be used as part of a reliable test of our hypothesis.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

Introduction 000	Negative argument 00●00	Positive Argument	Recapitulation 0000	References
An exam	ple: pendulum	model		

Suppose we want to use a pendulum to as part of a test of a hypothesis about acceleration, and we model the relationship between measurable quantities of the pendulum and acceleration with

$$I = \frac{aT^2}{4\pi^2}$$

where T is the period and I the length of the string.

Strictly speaking, however, the model has false assumptions.

Introduction 000	Negative argument 000●0	Positive Argument	Recapitulation 0000	References
Implication				

What does this show?

Just that a model having true assumptions are not necessary for having high confidence in the model result.

What matters instead is the reliability or trustworthiness of the model (in the given context!).

Introduction 000	Negative argument 0000●	Positive Argument	Recapitulation 0000	References
The poir	it vis-a-vis rohi	istness		

But if the truth of the model assumptions doesn't (directly) determine my confidence in the result, then inconsistency across model assumptions is largely irrelevant.

What matters instead is mutual unreliability.

But mutual inconsistency is not sufficient for mutual unreliability.

Hence the argument for disunity fails.

Introduction 000	Negative argument	Positive Argument	Recapitulation 0000	References

Positive Argument

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

Introduction 000	Negative argument	Positive Argument ●000000	Recapitulation 0000	References
The positi	ve argument			

Simply: given two plausible assumptions, I can show that there's a formal equivalence between (some?) cases of experimental and modeling robustness.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

Introduction	Negative argument	Positive Argument	Recapitulation	References
000	00000	0●00000	0000	
First assum	ption			

The one just argued for: what matters for our confidence in a proposition is the reliability of the model, not whether it has true assumptions.

Why does this matter?

There's a standard model of the value of variation across models owed to Woodward (2006).

But this model assumes that the success of one model rules out the success of the others—I need to drop that.



Let r_i indicate some important condition on the reliability of the model. Oversimplifying:



▲ロト ▲帰ト ▲ヨト ▲ヨト 三日 - の々ぐ

Introduction 000	Negative argument	Positive Argument 000●000	Recapitulation 0000	References
The second	assumption			

"More data" and "varied data" are conceptually distinct and can come apart.

(And therefore, obviously, if we're aiming to give an analysis of the second, we should try and distinguish it from the first.)

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

Introduction 000	Negative argument	Positive Argument	Recapitulation 0000	References
Distinguishi	ng varied data	from more dat	ta	

Imagine the data-collection limit, where we have infinite data from an experiment

- 1. No value from simply repeating the same test (more data).
- 2. Still some value to varying experiment (some data)

Why? Because the latter corrects for systematic errors in a way the first one doesn't.

Introduction 000	Negative argument	Positive Argument 00000€0	Recapitulation 0000	References
The picture	e			

Let r_i indicate some important condition on the reliability of the experiment



◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Introduction 000	Negative argument	Positive Argument 000000●	Recapitulation 0000	References
The lesson				

Given the two assumptions, there's a formal equivalence between robustness in the two cases.

◆□▶ ◆□▶ ◆三▶ ◆三▶ 三三 のへぐ

Strong evidence that unity is true (and disunity false) but not necessarily definitive.

Introduction	Negative argument	Positive Argument	Recapitulation	References

Recapitulation

◆□▶ ◆□▶ ◆臣▶ ◆臣▶ 臣 のへぐ

I've argued that robustness is a single, unified, epistemic phenomenon: there's no difference between robustness in modeling and experimental contexts.

But there is *something* to the intuition that robustness is more valuable in experimental contexts.

This something is a tendency—but probably an overemphasized one.

< □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > < □ > <

Introduction 000	Negative argument	Positive Argument	Recapitulation 0●00	References
Reasons for	r the tendency			

Speaking generally, the more knowledge we have a phenomenon—the better any one model that we can construct—the harder it is to build multiple (semi-independent) models.

By contrast, the more knowledge we have a phenomenon, the better (and more diverse) our experimental access.

Since the value of robustness depends in part on the quality of the tools involved, the implication is that there will tend to be more value to robustness in experimental cases, as those are the cases in which our tools are better.

Introduction 000	Negative argument	Positive Argument	Recapitulation 00●0	References
But also pr	obably overem	phasized		

That said, this tendency is probably overemphasized. The literature likes to compare exemplary experimental cases (e.g., Perrin) to cases in which the models employed are all very similar and have very little empirical evidence to support them.

But since robustness depends on those factors, the result is a sort of "no true scotsman": of course robustness across models doesn't provide good evidence if matching Perrin is the criteria of "good evidence."

When we insist on a *ceteris paribus* comparison—as I do in $\S2$ —the two look exactly the same.

Introduction 000	Negative argument 00000	Positive Argument	Recapitulation 000●	References
Conclusion				

I've done three things in this talk:

1 argued that the standard argument for disunity fails: we cannot move from mutual inconsistency of model assumption to a needing different analyses

- §2 provided a positive argument for unity on the basis of two plausible assumptions.
- $\S3$ recovered the original intuition.

Introduction	Negative argument	Positive Argument	Recapitulation	References



Cartwright, Nancy (1991). Replicability, Reproducibility, and Robustness: Comments on Harry Collins. *History of Political Economy* 23.1: 143–55.

Woodward, James (2006). Some Varieties of Robustness. Journal of Economic Methodology 13.2: 219–40.