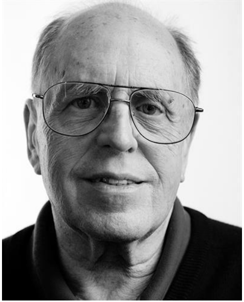
Stability in Climate Change Attribution

Corey Dethier

Minnesota Center for Philosophy of Science
University of Minnesota, Twin Cities

Nov. 16, 2024

George E. Smith (1938-2024)



This paper was originally intended for a symposium on George's work.

In essence, I'm applying the account of 'theory-mediated measurement' George develops in Smith (2014) and Smith and Seth (2020) to the measurement of the human contribution to warming.

An alternative title was "Can Attribution Science Close the Loop?"

Two claims

Claim 1: internal variability is not a particularly serious problem for measuring the human contribution to climate change.

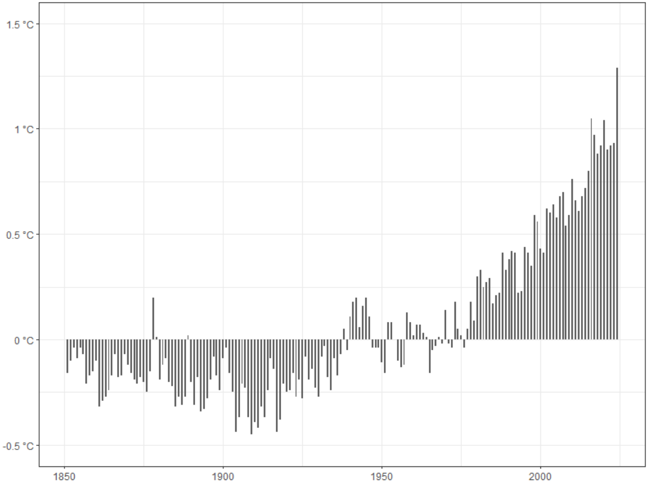
Claim 2: there is only one "logic" of confirmation by way of stability / robustness.

The plan

1. Internal variability and measuring the human contribution.
2. Stability and confirmation in general
3. Stability and confirmation in attribution studies.
4. The one logic of confirmation by way of stability.

Internal Variability and Climate Change Attribution

Climate change



Average global temperatures from May 1851 to April 2024. Data from NCEI.

What we knew by 1990 (if not much earlier)

1. Average global temperatures are going up.
2. Increasing CO_2 causes temperatures to go up, *ceteris paribus*.
3. Human actions have caused a (massive) increase in CO_2.

What we knew by 1990 (if not much earlier)

1. Average global temperatures are going up.
2. Increasing CO_2 causes temperatures to go up, *ceteris paribus*.
3. Human actions have caused a (massive) increase in CO_2.

---

4. (Probably) human actions are responsible for increasing temperatures.

Measuring the human contribution to warming

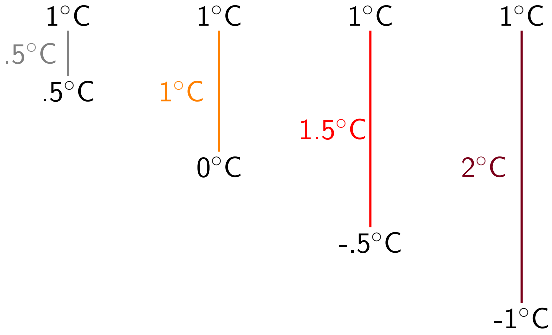
To **measure** the human contribution to warming, we need to know the state that the system would have exhibited without human intervention.

For discussion of attribution qua measurement process, see Dethier (2022).

See also Smith (2014) and Smith and Seth (2020) on the counterfactual nature of this condition.

Illustration

What's observed



What would have happened without human actions

Internal variability

In climate science:

What would have happened without human actions

=

"**internal variability**" **(IV)**.

A problem

IV can't be derived from first principles.

IV isn't observable: there's no version of the earth unaffected by climate change.

We have to rely on idealized simulations and risky extrapolations from paleoclimate analogues

A problem

IV can't be derived from first principles.

IV isn't observable: there's no version of the earth unaffected by climate change.

We have to rely on idealized simulations and risky extrapolations from paleoclimate analogues

$\Rightarrow$ uncertainty about **IV**

$\Rightarrow$ uncertainty about the human contribution.

Enter stability

Both Parker (2010) and Katzav (2013) identify **IV** as a serious problem in measuring the human contribution.

Parker (2010, 1090–91) suggests a remedy: *if* the measure of the human contribution is stable, we *might* have grounds for thinking that our estimates of **IV** are accurate.

As of 2010, different measures didn't exhibit much stability.

Things have changed!

	1986-2005	1995-2014	2006-2015	2010-2019
Observed	.69 (.52-.82)	.86 (.67-.98)	.94 (.76-1.08)	1.06 (.88-1.21)
Gillett et al. (2021)	.63 (.32-.94)	.84 (.63-1.06)	.98 (.74-1.22)	1.11 (.92-1.30)
Haustein et al. (2017)	.73 (.58-.82)	.88 (.75-.98)	.98 (.87-1.10)	1.06 (.94-1.22)
Ribes et al. (2021)	.65 (.52-.77)	.82 (.69-.94)	.94 (.80-1.08)	1.03 (.89-1.17)

The °C change in temperature relative to the period 1850-1900. The first row is the observed change (IPCC 2021, 320). The other rows are estimates for the warming attributable to humans (IPCC 2021, 442).

Questions

1. Under what conditions would stable results confirm our estimate of **IV**?
2. Do extant studies meet those conditions?
3. What does this mean for measuring the human contribution?

When stability is evidence

An astronomical example

Consider measuring the mass of the sun by way of the motions of a planet:

$$F_G \approx m_x a_x \qquad F_G = G m_x m_s r_{sx}^{-2}$$

$\searrow \qquad \swarrow$

An astronomical example

Consider measuring the mass of the sun by way of the motions of a planet:

$$F_G \approx m_x a_x \qquad\qquad F_G = G m_x m_s r_{sx}^{-2}$$

$$\searrow \qquad \swarrow$$

$$m_s \approx G^{-1} a_x r_{sx}^2$$

An astronomical example

Consider measuring the mass of the sun by way of the motions of a planet:

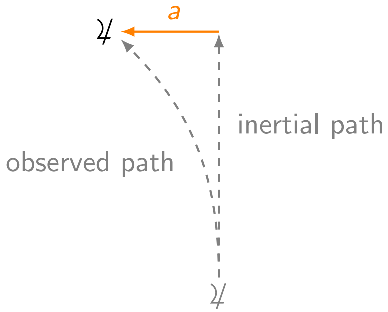
$$F_G \approx m_x a_x \qquad\qquad F_G = G m_x m_s r_{sx}^{-2}$$

$$\searrow \qquad \swarrow$$

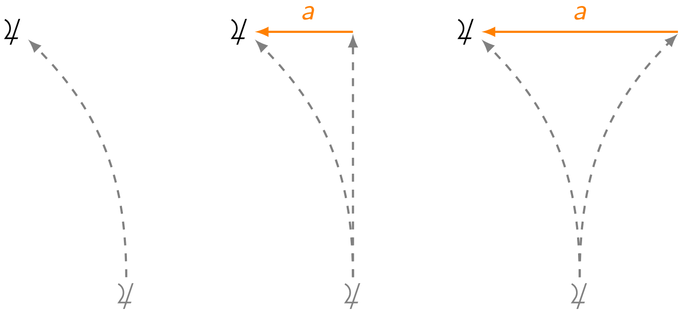
$$m_s \approx G^{-1} a_x r_{sx}^2$$

Problem: the result relies on assumptions – such as the principle of inertia – that we have no independent way of verifying.

Illustration



Illustration



How stability helps

If the principle of inertia is accurate, our measurements should be stable when we vary the planet in question.

That is: the m_s term in

$$m_s \approx G^{-1} a_x r_{sx}^2$$

should take on (approximately) the same value regardless of which planet we plug in.

What if our assumptions are false?

If the principle of inertia is inaccurate, our measurements should **not** be stable when we vary the planet in question.

After all, if

$$a_x \not\approx F_G / m_x$$

then

$$m_s \not\approx G^{-1} a_x r_{sx}^2$$

And thus it would be a massive coincidence if multiplying the wildly varying values of r_{sx}^2 by a_x together yielded to get stable m_s values.

The upshot

Summarizing:

1. The principle of inertia is accurate ⇒ stability in m_s.
2. The principle of inertia is inaccurate ⇒ instability in m_s.

If m_s is stable, we have good reason to believe that that the principle of inertia is accurate.

Smith (2014) and Smith and Seth (2020) refer to measurements satisfying these conditions as "theory-mediated."

Stability and internal variability

Extrapolating

Stable measures of the human contribution confirm our estimate of **IV** if:

1. Estimate of **IV** is accurate $\Rightarrow$ stable measure of the human contribution.
2. Estimate of **IV** is inaccurate $\Rightarrow$ unstable measure of the human contribution.

Extrapolating

Stable measures of the human contribution confirm our estimate of **IV** if:

1. Estimate of **IV** is accurate $\Rightarrow$ stable measure of the human contribution.
2. Estimate of **IV** is inaccurate $\Rightarrow$ unstable measure of the human contribution.

Essentially: we need the measure of the human contribution to be sensitive to realistic differences in the estimate of **IV**.

If the assumptions are false ...

The three studies estimate and use **IV** in different ways:

1. Gillett et al. (2021) use a CMIP6-based estimate as a filter to isolate that part of the data in which the signal is to be identified á la the classical method of Hasselmann (1993).
2. Ribes et al. (2021) use a contrasting Bayesian method; internal variability enters in during updating and they estimate it using a combination of prior research and CMIP6 data.
3. Haustein et al. (2017) use a CMIP5-based estimate and (so far as I can tell) only use it in generating uncertainty bands.

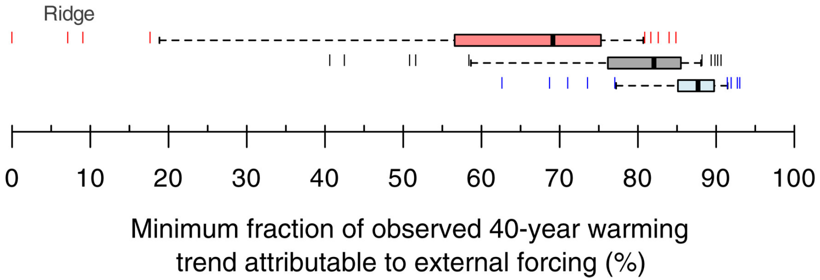
Sippel et al. (2021)

Currently, our best methods estimate that humans are responsible for a *minimum* of $80-90\%$ of observed warming.

How sensitive is this number to different estimates of **IV**?

See also Imbers et al. (2013, 2014).

Not very



From Sippel et al. (2021, Fig. 6).

Bad news for estimating internal variability

Upshot: the measure is **not** sensitive to (realistic errors in) the estimate of **IV**.

Which means:

> **Not**: Estimate of **IV** is inaccurate $\Rightarrow$ unstable measure of the human contribution.

But ...

What does this mean for measuring the human contribution?

Notice:

> **if** there is an error in measure of human contribution **because** of an error in **IV**

then we would expect Sippel et al. (2021) to find sensitivity, or:

> small differences in **IV** $\Rightarrow$ unstable results.

Good news for measuring the human contribution

Both of the following conditionals hold:

1. **IV** is not a problem for the measure of the human contribution ⇒ stability in the results of Sippel et al. (2021).
2. **IV** is a problem for the measure of the human contribution ⇒ instability in the results of Sippel et al. (2021).

Upshot

The stability of measures of the human contribution doesn't confirm our estimate of **IV**.

BUT, it does suggests that errors in **IV** are not likely to cause errors in our measurement.

There is only one logic of stability / robustness

Cases

I've reviewed three cases of reasoning from stability – what's sometimes called "robustness reasoning":

1. stable mass estimates → accuracy of the principle of inertia
2. stable attribution results → accuracy of the estimate of **IV**
3. stable results from Sippel et al. (2021) → accuracy of the measure of human contribution

(In the paper, I survey five additional examples.)

The logic of stability

All of these examples obey the same "logic":

Stability in X confirms a hypothesis H if:

1. H predicts stability in X; and
2. $\neg$H predicts instability in X.

Individual cases differ (dramatically!) according to whether – and to what degree – these two conditions are satisfied.

Insofar as they differ in other ways, those differences aren't relevant to confirmation – hence one "logic."

For details of how to capture this idea in a Bayesian formalism, see Dethier (2024a,b) or Myrvold (1996, 2017).

Two lessons

Lesson 1: internal variability is not a particularly serious problem for measuring the human contribution to climate change.

Lesson 2: there is only one "logic" of confirmation by way of stability / robustness.

End

Thank you!!

Dethier, Corey (2022). Calibrating Statistical Tools: Improving the Measure of Humanity's Influence on the Climate. *Studies in the History and Philosophy of Science* 94: 158–66. DOI: 10.1016/j.shpsa.2022.06.010.

— (2024a). Contrast Classes and Agreement in Climate Modeling. *European Journal for Philosophy of Science* 14.14: 1–19. DOI: 10.1007/s13194-024-00577-6.

— (2024b). The Unity of Robustness: Why Agreement Across Model Reports is Just as Valuable as Agreement Among Experiments. *Erkenntnis* 89: 2733–52. DOI: 10.1007/s10670-022-00649-0.

Gillett, Nathan P. et al. (2021). Constraining Human Contributions to Observed Warming since the Pre-industrial Period. *Nature Climate Change* 11: 207–12. DOI: 10.1038/s41558-020-00965-9.

Hasselmann, Klaus (1993). Optimal Fingerprints for the Detection of Time-dependent Climate Change. *Journal of Climate* 6.10: 1957–71. DOI: 10.1175/1520-0442(1993)006<1957:OFFTDO>2.0.CO;2.

Haustein, Karsten et al. (2017). A real-time Global Warming Index. *Scientific Reports* 7 (15417): 1–6. DOI: 10.1038/s41598-017-14828-5.

Imbers, Jara et al. (2013). Testing the Robustness of the Anthropogenic Climate Change Detection Statements using Different Empirical Models. *Journal of Geophysical Research: Atmospheres* 118.8: 3192–99. DOI: 10.1002/jgrd.50296.

Imbers, Jara et al. (2014). Sensitivity of Climate Change Detection and Attribution to the Characterization of Internal Climate Variability. *Journal of Climate* 27.10: 3477–91. DOI: 10.1175/JCLI-D-12-00622.1.

IPCC (2021). *Climate Change 2021: The Physical Science Basis*. Ed. by Valérie Masson-Delmotte et al. Sixth Assessment Report of the Intergovernmental Panel on Climate Change. Cambridge: Cambridge University Press.

Katzav, Joel (2013). Severe Testing of Climate Change Hypotheses. *Studies in History and Philosophy of Science Part B* 44.4: 433–41. DOI: 10.1016/j.shpsb.2013.09.003.

Myrvold, Wayne (1996). Bayesianism and Diverse Evidence: A Reply to Andrew Wayne. *Philosophy of Science* 63.4: 661–65. DOI: 10.1086/289983.

— (2017). On the Evidential Import of Unification. *Philosophy of Science* 84.1: 92–114.

Parker, Wendy S. (2010). Comparative Process Tracing and Climate Change Fingerprints. *Philosophy of Science* 77.5: 1083–95. DOI: 10.1086/656814.

Ribes, Aurélien, Saïd Qasmi, and Nathan P. Gillett (2021). Making Climate Projections Conditional on Historical Observations. *Science Advances* 7.4: 1–9. DOI: 10.1126/sciadv.abc0671.

Sippel, Sebastian et al. (2021). Robust Detection of Forced Warming in the Presence of Potentially Large Climate Variability. *Science Advances* 7.43: 1–17. DOI: 10.1126/sciadv.abh4429.

Smith, George E. (2014). Closing the Loop: Testing Newtonian Gravity, Then and Now. In: *Newton and Empiricism*. Ed. by Zvi Beiner and Eric Schliesser. Oxford: Oxford University Press: 262–351.

Smith, George E. and Raghav Seth (2020). *Brownian Motion and Molecular Reality: A Study in Theory-Mediated Measurement*. Oxford: Oxford University Press.