

Experiments as Performance Art

no IRBs necessary!

Kevin Munger

Date: Sep 15, 2025

URL: <https://kevinmunger.substack.com/p/experiments-as-performance-art>

Scientific progress is not linear but dialectical. The current “credibility revolution” in social science came in response to longstanding concerns about the reliability of empirical research, particularly regarding causal inference. For decades, much of social science relied on observational data and complex statistical models, which produced results that were difficult to interpret and which made it all too easy to confuse correlation and causation.

This credibility revolution made progress by insisting on “credible” causal inference. Central to this movement was the widespread adoption of randomized controlled trials (RCTs), natural experiments, and other quasi-experimental methods that prioritized internal validity.

Anyone reading a social science blog today is familiar with RCTs. They’re widely (but not universally) believed to be the “gold standard” research method, and while I am one of the critics of this label, today I advance an even more radical claim about RCTs: not only are they ***not*** the gold standard, they’re ***not even a research method***.

At least not, as we shall see, in one specific but extremely important sense.

Dialectical scientific progress rolls on. Where we once had a crisis of internal validity (the impetus for the credibility revolution), we now have a crisis of *external* validity. We can conduct an essentially perfect RCT in one time and place; we can be certain that the treatment group and the control group have different outcomes, and that this effect was definitely caused by the treatment and the treatment alone. But this effect cannot be generalized to other times and places. This perfect golden knowledge nugget cannot yet be melted down with other knowledge nuggets. No human mind or machine learning algorithm can forge these nuggets into a holy grail: the ability to reliably predict (with anything near the level of precision of the original studies) what will happen when the treatment from the RCT is deployed again, in the real world.

Statisticians and practicing social scientists are valiantly applying their tools to this problem, but there are many issues. One of the biggest hurdles to conducting large-scale field experiments is ethical: people think, for good reason, that mad social scientists shouldn’t be able to run around conducting experiments on society. At a minimum, there should be guidelines: when possible, make sure to get informed consent from research subjects; don’t put subjects at risk of unnecessary harm; if trialling something that is expected to cause benefits (like a conditional cash transfer experiment), be sure to give the same benefits to the subjects in the control group after the experiment ends; don’t use AI to make up stories about being a rape

victim in order to change people's view on Reddit; but most importantly, ***make sure to get IRB approval.***

Institutional Review Boards (IRBs) are the university-level entities tasked with ensuring the ethics of "human subjects research."¹ They were standardized by the 1991 "Common Rule", updated in 2018 to expedite some of the cumbersome approval process. As this is an official legal document, these terms have precise and exhaustively defined meanings. Much of the document is devoted to defining "human subjects"; as anyone who has submitted something to an IRB knows, there are different levels of protections and restrictions involved with working with vulnerable populations (children, prisoners) and for more intrusive interventions (testing a new drug versus asking a survey question). There are, unfortunately, tragic historical examples for each line of this definition.

Less attention is given to the remainder of the phrase "human subjects research." "Research" here is defined as a 'systematic investigation...designed to develop or contribute to generalizable knowledge' – per section 46.102(i).

This seem innocuous – especially from the perspective of RCTs as the gold standard method. But given what we now know about the limits of external validity, we can't be sure. According to my epistemic commitments, the results of RCTs are not generalizable. They are, therefore, ***not research***, and are thus exempt from IRB approval.²



image unrelated

This would be tendentious trolling if I didn't believe it. If RCTs aren't research, what are they? Should we do them? Why?

The value of RCTs comes not from the control group but from the treatment group -- from the **action**. They are best understood as a formal kind of performance art. Each society enjoys performances in the idiom of its respective culture. Martial societies find meaning in ritual combat; religious societies in pious displays of devotion or spiritual rapture.

Our society is scientific, even scientistic. We appreciate the performance of scientific rituals, big data crunching and demonstrations of control over nature or our fellow citizens. Social science experiments don't provide us with airtight guarantees about what will happen in the future -- and it's just as well that they don't, because such guarantees are incompatible with democracy and human freedom. What these experiments provide us is the ability to see the social world in a different way.

In a [review of the Meta2020 academic partnership](#), I wrote that

My aesthetic appreciation for these experiments cannot be overstated. They are simply beautiful. Thinking as I sometimes do of social science experiments as performance art, the craft and the vision on display is deeply satisfying.

I find this perspective liberating. "[you can just do stuff](#)" has become a rallying cry for the Silicon Valley anti-progressives, who feel constrained by both social norms and government regulations. On this point I agree with them: we need to recover the spirit of individual initiative outside of institutional constraint. The "founder mode"/start-up/a16z impulse is central to the [American ethos](#), and it is genuinely a political problem that this impulse is at present finds its most fertile soil in AI and defense startups but seems to wither in government, universities and [general intellectual culture](#).

One of the key battlegrounds in the emerging ideological split between the center-left credentialed establishment and the Silicon Valley right is about the role of *science*. Each side is explicitly in favor of science, but they differ in their conception of what *science* means.

The SV people have explicitly embraced the scientist as a radical truth-teller and handmaiden to technological progress, a cross between Galileo tweeting "*E pur si muove!*" and von Neumann inventing game theory and then confidently telling Truman that game theory says we have to genocide the Russians with nukes. In contrast, the credentialed establishment tells us to "trust the science" as if this were a static, eternally valid scripture rather than a [process](#) for developing expertise and refining our beliefs.

The bureaucratization of "ethics" represented by the IRB certainly moves academic science in the latter direction. Scott Alexander has a [shocking story of his experience](#) with a medical IRB, and while I haven't experienced anything so dramatic, the scientific-entrepreneurial experience is significantly dimmed by spending six months going back and forth explaining that, say, Twitch users are pseudonymous already and that asking them to give me their email address so they can subsequently opt out their data being used in an experiment is in fact

introducing more risk of de-anonymization than simply letting this bureaucratic box go un-checked.

To re-democratize science, we need to encourage everyone to believe that they can just do stuff — that they can and should try to understand how to improve their local situation by trying things out. That is, that you can *literally* Do Your Own Research. The idea that only a very specific subset of human actions contributes to “generalizable knowledge,” and that this formally-defined **research** is something defined, controlled, and gatekept by government-aligned academics, enervates the non-specialist and (in the internet era) inevitably fuels anti-establishment backlash.

Tech startups are allowed and financially encouraged to run uncontrolled “experiments” on hundreds of millions of people. They “try things out” all the time; social media was a massive, poorly-designed experiment, and now they’re just letting chatbots rip throughout the most alienated and lonely members of society. They don’t aspire to “generalizable knowledge,” so they aren’t doing “human subjects research,” so no ethics review is necessary.

The message of the IRB is that the ethical risk of an RCT comes from the **addition of a control group** to the nihilistic rollout of a massive new social technology. Personally, I disagree that RCTs generate “generalizable knowledge” to a categorically different degree than does other forms of systematic inquiry. This ethico-methodological straightjacket is bad for science and bad for society. The Silicon Valley message is nominally more free-spirited but in practice produces an equally restricted science aimed only at maximizing Monthly Active Users in service of shareholder value.

I’ll give the final word to John Dewey:

The future of our civilisation depends upon the widening spread and deepening hold of the scientific habit of mind.

1. Note: IRBs are specific to the US, at least as the following regulations are concerned. Many universities in other countries have adopted something similar but the language may of course differ.

2. Thanks to P Aronow for first pointing this contradiction out to me.