Evidence-based Decision Making
Experiments

Rui Mata, FS 2023

Version: April 17th, 2023

Die Fakultät für Psychologie der Universität Basel lädt Sie ein!

DIENSTAG, 9. MAI 2023, 17:00

INFORMATIONSVERANSTALTUNG

ZUM MASTERSTUDIUM IN SOZIAL-, WIRTSCHAFTS- UND ENTSCHEIDUNGSPSYCHOLOGIE

17:00 Uhr
FAKULTÄT FÜR PSYCHOLOGIE
MISSIONSSTRASSE 62A
HÖRSAAL 00.006

DIENSTAG, 9. MAI 2023, 17:30

PSYCHOLOGIE IN DER PRAXIS

ABSOLVENTEN/INNEN DER MASTERVERTIEFUNGSRICHTUNG SOZIAL-, WIRTSCHAFTS- UND ENTSCHEIDUNGSPSYCHOLOGIE BERICHTEN VON IHREN BERUFSERFAHRUNGEN NACH DEM STUDIUM

MIT ANSCHLIESSENDEM APÉRO

17:30 Uhr
FAKULTÄT FÜR PSYCHOLOGIE
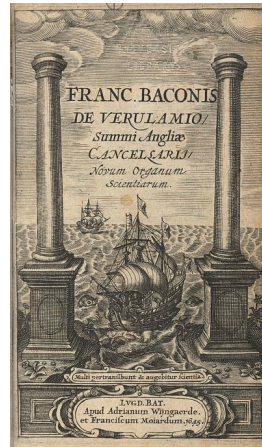**MISSIONSSTRASSE 62A**
HÖRSAAL 00.006

Goals

- Understand the nature of causal inference as the comparison of treatment to some counterfactual

- Understand that RCTs/experiments have desirable properties for causal inference – but also have limitations…

Evidence-based decision making

Francis Bacon
(1561-1626)

1620

Varian, H. R. (2016). Causal inference in economics and marketing. Proceedings of the National Academy of Sciences of the United States of America, 113(27), 7310–7315. http://doi.org/10.1073/pnas.1510479113

Bacon suggests that one can draw up a list of all things in which the phenomenon to explain occurs, as well as a list of things in which it does not occur. Then one can rank the lists according to the degree in which the phenomenon occurs in each one. Then one should be able to deduce what factors match the occurrence of the phenomenon in one list and do not occur in the other list, and also what factors change in accordance with the way the data had been ranked.

"The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment"

The gold standard…

Experiments/Randomised control trials (RCT)

A type of scientific experiment, where the people being studied are randomly allocated one or other of the different treatments under study. RCTs are considered the gold standard for a clinical trial. RCTs are often used to test the efficacy or effectiveness of various types of medical intervention and may provide information about adverse effects, such as drug reactions. Random assignment of intervention is done after subjects have been assessed for eligibility and recruited, but before the intervention to be studied begins.

$$Y = B_0 + B_1 group$$
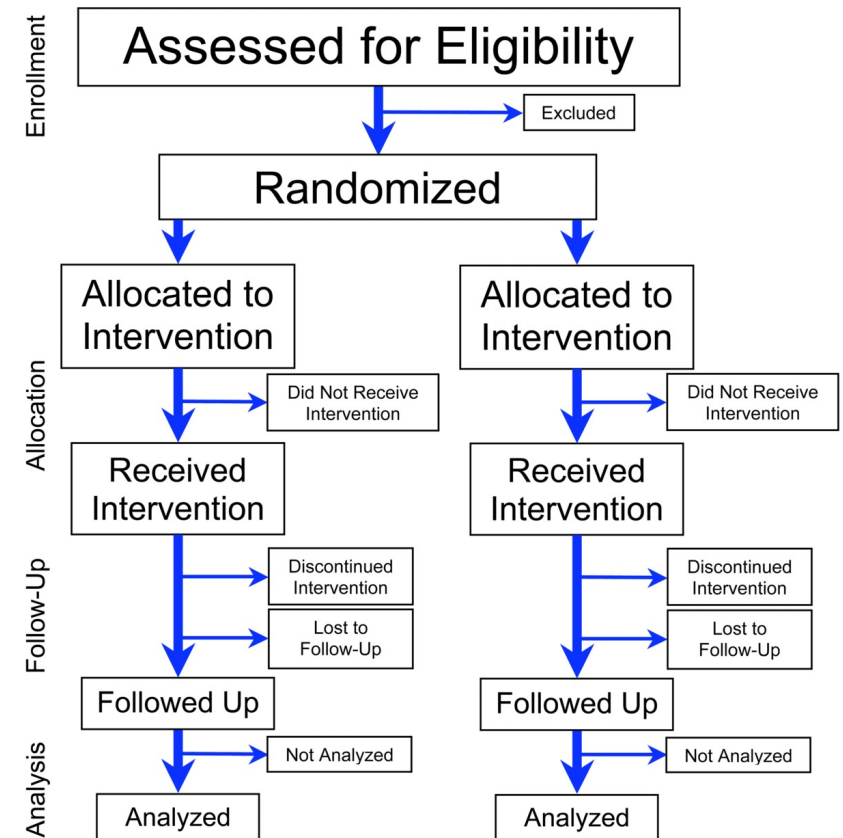
Shorter, E. (2011). A brief history of placebos and clinical trials in psychiatry. *Canadian Journal of Psychiatry, 56*(4), 193–197.

The gold standard…

Experiments/Randomised control trials (RCT)

http://www.consort-statement.org

The Salk Polio Vaccine Trial & the Cutter Incident

• The 1954 Salk Polio vaccine trial was the largest RCT (a double-blind, randomized, and placebo-controlled study) ever conducted, involving over 1.8 million children, to test the safety and efficacy of a polio vaccine developed by Jonas Salk.

• The results showed that the vaccine was safe and effective in preventing polio.

• In 1955, shortly after the Salk polio vaccine was licensed, a manufacturing error at one of 5 licensed laboratories, Cutter Laboratories, resulted in the contamination of some batches of the vaccine with live polio virus, which led to an outbreak that affected a few hundred children, including some deaths and cases of permanent paralysis, known as the Cutter incident.

• The Cutter incident led to significant changes in vaccine regulation including the creation of oversight agencies and legislation.

The Cutter incident is an example of the problems that may arise from generalizing RCTs – and the continued need for evaluation…

A manufacturing error at Cutter Laboratories resulted in the contamination of some batches of the vaccine with live polio virus

Offit, P.A. (2005). The Cutter incident, 50 years later. *N Engl J Med*. 352, 1411-1412.

Dawson, L. (2004). The Salk polio vaccine trial of 1954: Risks, randomization and public involvement in research. *Clinical Trials, 1*, 122–130.

The gold standard is not always gold…

Experiments/Randomised control trials (RCT)

- **Efficacy vs. effectiveness**: Trials may not be widely applicable in real-world conditions….

- **Generalizability**: Results may not always generalize to other samples (e.g. inclusion /exclusion criteria)

- **Ethical limitations**: randomisation requires experimental equipoise: one cannot ethically randomise participants to some treatments (no-schooling condition)

On the horizon: Autonomous Scientific Agents
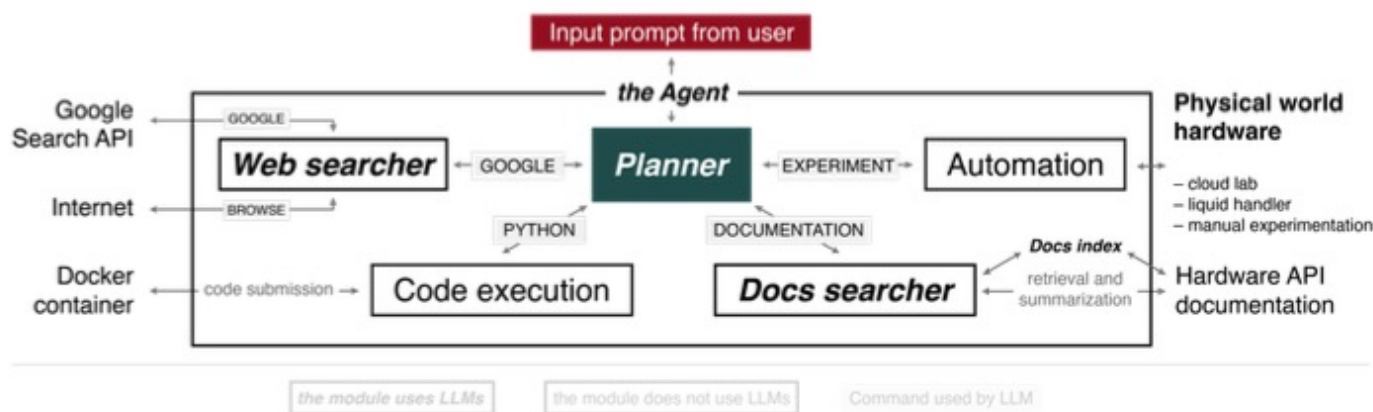
Figure 1. Overview of the system architecture. The `Agent` is composed of multiple modules that exchange messages. Some of them have access to APIs, the Internet, and Python interpreter.

In this paper, we presented an Intelligent `Agent` system capable of autonomously designing, planning, and executing complex scientific experiments. Our system demonstrates exceptional reasoning and experimental design capabilities, effectively addressing complex problems and generating high-quality code. However, the development of new machine learning systems and automated methods for conducting scientific experiments raises substantial concerns about the safety and potential dual use consequences, particularly in relation to the proliferation of illicit activities and security threats. By ensuring the ethical and responsible use of these powerful tools, we can continue to explore the vast potential of large language models in advancing scientific research while mitigating the risks associated with their misuse.

Boiko, D.A., MacKnight, R. & Gomes, G. L. (2023). Emergent autonomous scientific research capabilities of large language models. https://doi.org/10.48550/arXiv.2304.05332

Summary

"The critical step in any causal analysis is estimating the counterfactual—a prediction of what would have happened in the absence of the treatment"

RCTs are great **but** do not guarantee effectiveness, generalizability, or ethical treatment of participants…

Automation is on the rise – ethical and safety issues will be crucial!