

AFTER 160 YEARS: A PAUSE FOR REPLICATION?

INTERACTIVISM 2021

ROBERT L. CAMPBELL

CLEMSON UNIVERSITY (EMERITUS)

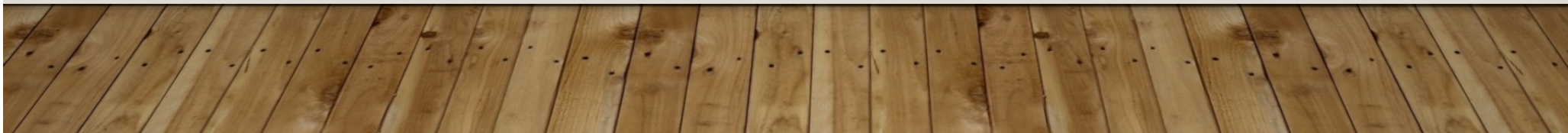
JUNE 26, 2021

SOME OF THOSE GATHERED HERE ARE PSYCHOLOGISTS

- Interactivist theory pertains to multiple disciplines, many of them represented in this conference
- All the same, interactivism originated in psychology
- Barring a mass rearrangement of Donald Campbell's fish scales
 - Much of the relevant empirical work that will test interactivist or related hypotheses on language, cognition, emotions, social interaction, and development, will be carried out by psychologists

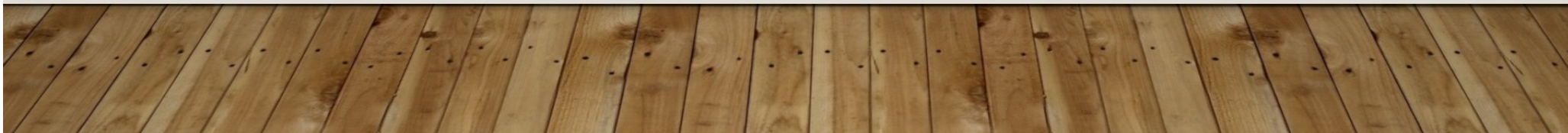
MODERN PSYCHOLOGY HAS BEEN COLLECTING DATA FOR A WHILE

- Wilhelm Wundt published *Contributions to a Theory of Sense Perception* in 1862
 - Wherein he announced his plan to establish experimental psychology
- Wundt's *Principles of Physiological Psychology* (first edition, 1873-1874)
- Growing numbers of trained psychologists have been conducting empirical studies ever since
- In too many areas of psychology, data sets keep accumulating, without enabling anybody to stand higher or see farther



WE NEED TO RETHINK THE WAY WE CONDUCT EMPIRICAL STUDIES AND THE CONCLUSIONS WE DRAW FROM THEM

- Psychological research faces other threats in today's world
 - Higher education in some developed countries has undergone a bubble, now in the slow phase of deflating
 - Critical Fill-in-the-Blank theory, if it takes control of psychology the as it has with some of the humanities, will be the end of psychology
- But while psychologists can afford to conduct empirical studies without ideological supervision, the thing to concentrate on is replication
- The manner in which psychologists conducted and published in the past has left us with too much that we thought we knew and have no particular reason to believe we really did



HYPOTHESIS TESTING FROM A LOGICAL POINT OF VIEW

- Can't be tested empirically: metaphysical hypotheses
 - Some psychologists have still tried...
- Can be tested empirically: universal hypotheses
- Can be tested empirically: existential hypotheses

UNIVERSAL HYPOTHESES

- About **all** members of a category
- “Every normal adult human being has a working memory”
- This may be true, but we can’t *verify* it
 - We’d need data on every normal adult human being alive today
 - Who else would we need data on?
- If it’s false, we can *falsify* it
 - One reliable counterexample will do

EXISTENTIAL HYPOTHESES

- About **some** instances of a category (at least one)
- “Some human beings can correctly enumerate 100+ objects very rapidly, without counting”
- Can you *verify* it, if it’s true?
 - One reliable example will do
- Can you *falsify* it, if it’s false?
 - Logically, falsifying it is equivalent to verifying that *no human being* can (could, will be able to) do what is claimed
 - Consequently, no

METAPHYSICAL HYPOTHESIS 1

- “For *every* operant behavior that a human being can add to his or her behavioral repertoire, *there is a* history of operant behavior and reinforcement*”
- *that conforms to the guidelines for schedules of reinforcement

METAPHYSICAL HYPOTHESIS 2

- *For every* problem that a human being can solve, *there is* a computer program that solves the problem the same way*.
- *i.e., carries out the same steps, makes the same errors as human beings (with roughly the same frequency), and does not make errors human beings are not observed to make

METAPHYSICAL HYPOTHESES ARE FAMILIAR TO PARTICIPANTS HERE

- Once in a while a metaphysical hypothesis is refuted by an *in-principle* argument
- A few in-principle arguments have been both sound and widely accepted by psychologists
 - Chomsky vs. Skinner
 - Minsky and Papert vs. 2-layer Perceptrons
- Other such arguments may not have been widely accepted by psychologists, but we know how to make a case for their soundness

HOW ABOUT THE UNIVERSAL AND EXISTENTIAL HYPOTHESES?

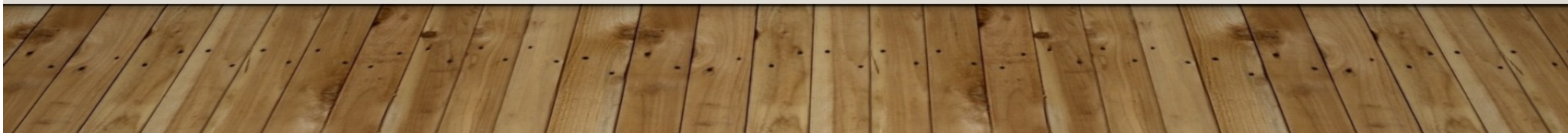
- Any good theory of anything needs lots of universal hypotheses
- Empirical research in psychology has a lousy track record with universal hypotheses
- *Logically, no finite set of results from empirical studies can verify a universal hypothesis*

REPLICATION IS LOGICALLY REQUIRED FOR TESTS OF UNIVERSAL HYPOTHESES

- As Karl Popper noted, the only thing researchers can do with universal hypotheses is *keep testing them*
- Keep trying to knock them down, see whether they keep standing
- Hence, nothing will work except replicating, replicating, replicating

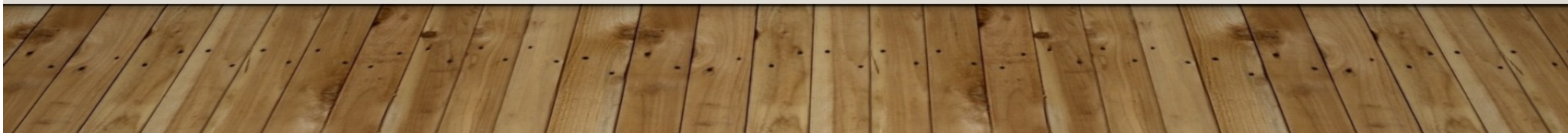
PSYCHOLOGISTS HAVE A LONG HISTORY OF DISDAIN FOR REPLICATION

- Psychology journals that publish empirical work (i.e., nearly all of them) until very recently weren't interested in publishing replication studies
- *One article that makes a splash* with data supporting a universal hypothesis has often made far more impact than a long series of reports of well-conducted studies trying to falsify that hypothesis (and perhaps succeeding)
- Worse: if the replication studies reported no statistically significant results, they were typically rejected for publication

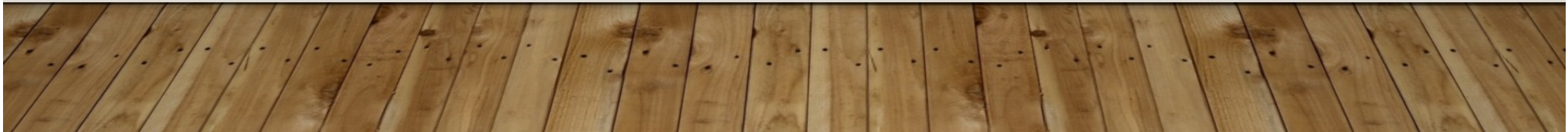


THE STUDY THAT MAKES A SPLASH AND THAT NO ONE BOTHERS TO REPLICATE (PROBABLY SHOULDN'T ATTEMPT TO REPLICATE)

- A single study (or set of studies) can have a lasting influence on the field even though
 - No one has tried to replicate it or them
 - Some have tried to replicate them, but the results of efforts to replicated are subsequently ignored
- In the worst case, the study isn't *worth* replicating because it was conducted so badly
- Or the study isn't worth replicating because it was conducted so unethically



AN EXPERIMENT NOT WORTH REPLICATING

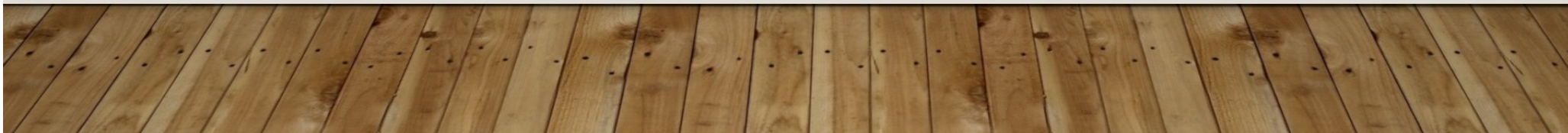


THE STANFORD PRISON EXPERIMENT (DATA COLLECTED 1971)

- This is usually taken as a prime case of research conducted unethically
- But Zimbardo et al.'s experiment violated the ethics of working with human subjects in multiple ways
 - All we usually hear about is the mistreatment of those assigned to be prisoners
- It was also reported in ways that violated the ethics of scientific communication
 - The misreporting is scarcely ever mentioned

THE STANFORD PRISON EXPERIMENT

- No one would want to conduct a straight replication of this experiment
- It would be improper even if a replication could be done without mistreating any participants and getting informed consent before they participated
- It would be improper to do so even without mentioning in recruitment advertisements that it is “a study of prison life”
- There is no merit in recapitulating inadequate data collection and multiple episodes of interference by members of the research team while the experiment was being run



DOWNSTREAM EFFECTS OF AN EXPERIMENT THAT NO ONE SHOULD REPLICATE?

- Harm to the reputation of the lead investigator?
- Discredit upon the main hypothesis of the experiment?

HISTOIRE D'UN MENSONGE



ENQUÊTE SUR L'EXPÉRIENCE DE STANFORD

Thibault Le Texier

Conduite en 1971 par le professeur Philip Zimbardo, l'«expérience de Stanford sur la prison» a vu vingt-deux étudiants volontaires jouer les rôles de gardiens et de prisonniers au sein d'une fausse prison installée dans l'université Stanford.

L'expérience devait durer deux semaines mais elle fut arrêtée au bout de six jours, résume Zimbardo, car «les gardiens se montrèrent brutaux et souvent sadiques et les prisonniers, après une tentative de rébellion, dociles et accommodants, même si la moitié d'entre eux furent si perturbés psychologiquement qu'ils durent être libérés plus tôt que prévu».

Devenue presque aussi célèbre que l'expérience de Stanley Milgram sur l'obéissance et souvent citée en exemple de l'influence des situations sur nos comportements, l'expérience de Stanford est pourtant plus proche du cinéma que de la science : ses conclusions ont été écrites à l'avance, son protocole n'avait rien de scientifique, son déroulement

ZONES

18 € (04.2018)
ISBN 978-2-35522-120-0



9 782355 221200

THE ETHICS OF SCIENTIFIC COMMUNICATION

- Issues:
 - Honest presentation of data
 - Honest presentation of sources and attribution of authorship
 - Minimizing the effects of bias and factional conflict in scientific research
- How has empirical research in psychology handled honest presentation of data?

DATA PRESENTATION

- Obviously, making up data that don't exist is unethical
- *Outright faking used* to be considered vanishingly rare in psychology
 - Sir Cyril Burt published with imaginary coauthors, but whether he faked data may never be known (the paper files were lost or destroyed long ago)

DATA PRESENTATION: FAKING OR ALTERING DATA SETS 1

- Diederik Stapel (Dutch social psychologist) was caught faking data in 2011
 - Some data were “merely” neatened up by altering or removing values from the data sets
 - Other “studies” turned out to have *no real data* at all
 - Journals great and small had published these items

DATA PRESENTATION: FAKING OR ALTERING DATA SETS 2

- Stapel had been altering data for 15 years, and outright faking for at least 10
- It was nearly impossible to get any academic administrator to call for an investigation of what he had been doing
- Fortunately, one official listened, the investigation was commissioned, and the Levelt Report is worth reading
- Stapel's memoir, *Ontsporing*, is unlikely to be published in translation

DATA PRESENTATION THAT IS MERELY MISLEADING

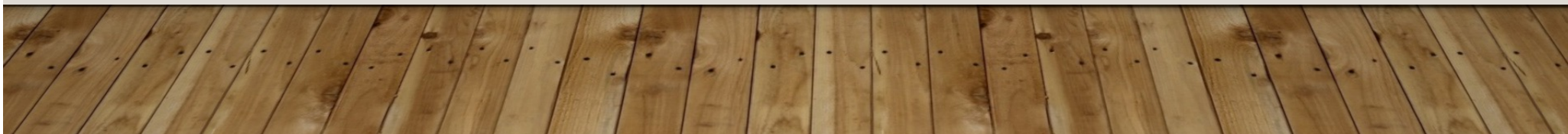
- *Misleading* presentation is more complex, because the norms of some specialties would be unacceptable in others
- Sometimes it's OK to average each individual participant's data over many trials, sometimes it's not
- Sometimes it's OK to exclude data from many individuals from the the final data set, sometimes it's not

QUESTIONABLE RESEARCH PRACTICES

- For roughly a decade, worries about replication in psychology (not just in the wake of fabricated or altered data) have led to a drive against “questionable research practices”
- QRPs include
 - Changing hypotheses after seeing the data
 - Excluding dependent variables from the data set after all data have been collected
 - Running frequent interim statistical analyses on the data and “optionally stopping” as soon as a key inferential statistic comes out significant
- All are good things to avoid, but
 - Traditional peer review practices aren’t going to catch most of them
 - The QRP portion of the blue-ribbon panel’s report on Stapel was the one part that drew a lot of public pushback from psychologists

COMBATTING QUESTIONABLE RESEARCH PRACTICES

- There's an initiative to *declare one's hypotheses before collecting data*—even to place a *research plan* in a public registry
 - I don't see this one working out (even stating how many participants one intends to recruit doesn't indicate how many actually will be)
- There's also push for putting *complete data sets* (anonymized when necessary) in *publicly accessible repositories*
 - In most cases, this is a good idea (though some informed consent procedures will have to be changed)
 - The effort required will be still be considerable, but other researchers will be able to take their look at the data set and the researcher's decisions about of summarizing and analyzing it



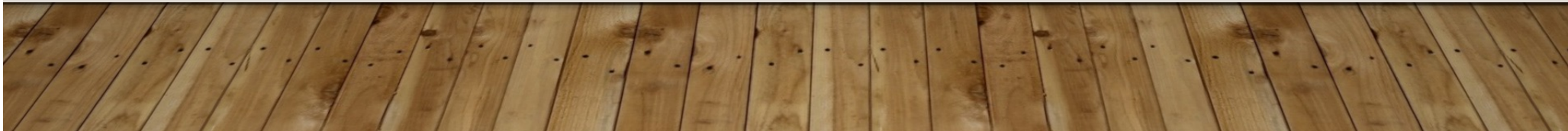
SOMEONE HAS TO CARRY OUT THE REPLICATIONS

- Damping down on QRPs will help with *replicability*
- Allowing examination and reanalysis of raw data will also enhance *replicability*
- But other researchers, preferably with different assumptions and biases and ideas about designing studies, *still have to run the replication studies*
- The journals *have to publish them*
- And the rest of the field still *has to take replication studies seriously*

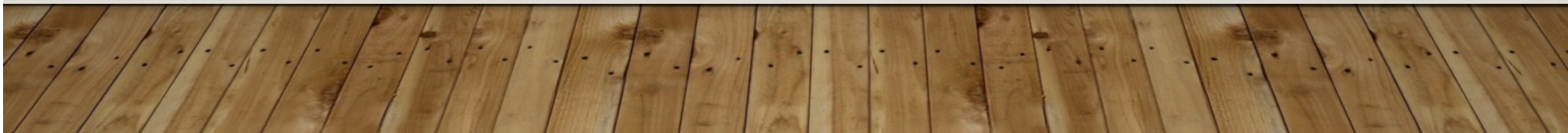
A FURTHER WRINKLE

- QRPs often involve the abuse of inferential statistics
- Optional stopping substitutes an ad hoc conditional distribution for the distribution that the test statistic is supposed to have
- Unfortunately, using inferential statistics “correctly” in psychology won’t heal all the ills of empirical research
- *How often do test statistics have the distribution they’re supposed to have?*

**WHAT KIND OF SAMPLING IS
ALWAYS ASSUMED IN STATISTICS?**

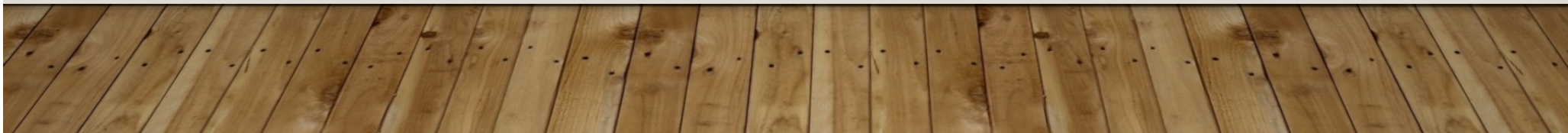


**WHAT KIND OF SAMPLING IS HARDLY
EVER USED IN PSYCHOLOGY?**



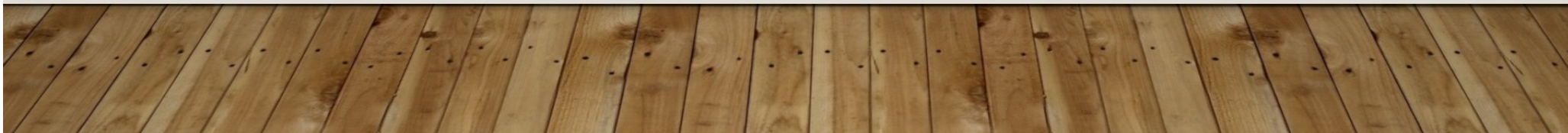
SAME ANSWER!

- The development of inferential statistics assumes *random* sampling from a population
- *Random* samples are the only samples that mathematicians know about
- Psychology and the other social sciences rarely draw *random* samples from the populations they are trying to generalize about
 - Whether it's a *really large* population, such as all normal adult human beings currently alive
 - Or a smaller population that researchers know more about, such as all registered voters in Pennsylvania in 2020



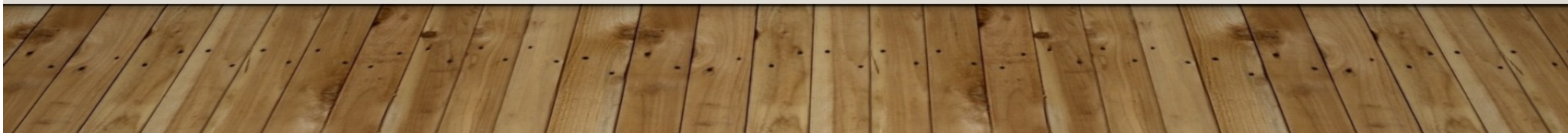
RANDOM SAMPLES

- A *random* sample is one in which *every instance, member, or individual in the population has an equal chance of being in the sample*



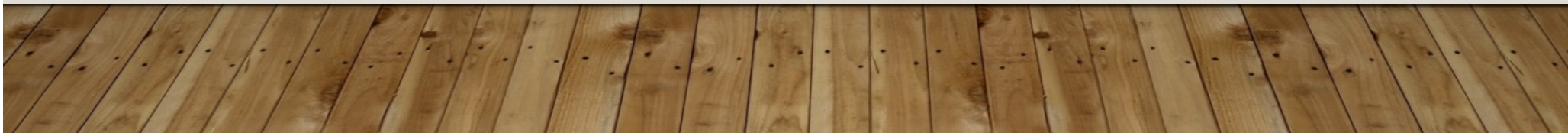
GETTING A RANDOM SAMPLE IS (USUALLY) A WELL-DEFINED PROBLEM

- One might naïvely expect psychologists to use a lot of random samples from their populations of interest
- Pick up an issue of any psychology journal that publishes empirical studies, and go through the Method section for each study reported in the issue
- It will be extraordinary if you find one study that used a random sample
- I won't recommend lighting a match and applying a Humean procedure to the rest, but...



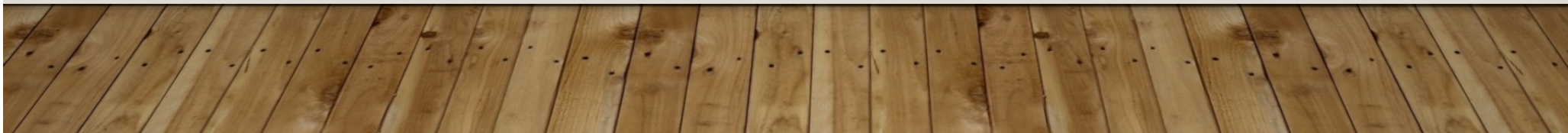
PSYCHOLOGISTS RARELY TRY TO GET RANDOM SAMPLES

- Suppose we want to generalize to all normal adult human beings
- *Some* members of this population aren't too hard to reach with our recruitment materials
- Others?
 - We still often don't know exactly who lives in some area
 - Does a person from the Parrot's Beak of Irian Jaya (western New Guinea) have the same chance of being in our sample as an individual from the Northern Territory of Australia?
 - Or we know that outsiders won't be allowed to recruit some people for a study
 - Does an individual from North Korea have the same chance of being in our sample as an individual from Taiwan?
- *Precisely because we know how what would be required to get a random sample from a large population, we know that we won't succeed at it*



PSYCHOLOGISTS RARELY TRY TO GET RANDOM SAMPLES

- With rare exceptions, participation in psychology studies ethically requires voluntary consent
- If we randomly select some person from our population, he or she may decline to participate
- *Perhaps* we'll try someone else from the population at random as a (possible) replacement
 - Except, what if those who don't want to participate differ in a bunch of important ways from those who do?



SMALLER POPULATIONS?

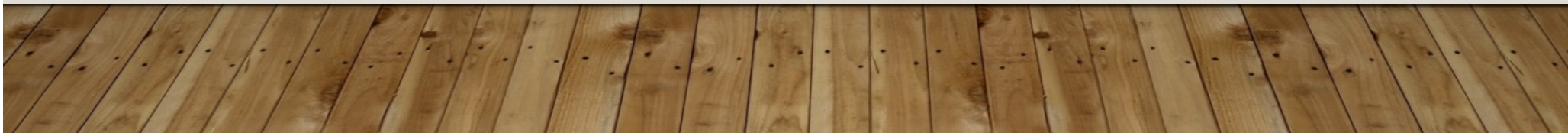
- With a small population, there are fewer barriers to getting a random sample
- For instance, if you wanted to draw a random sample of all registered voters in Pennsylvania in 2020, you could find out (within the limits of accuracy for the official records) exactly who was in the population
- Still, in most studies all participation has to be voluntary...

THE ELEPHANT IN THE ROOM

- What kind of sample do psychologists nearly always use?

CONVENIENCE SAMPLES

- Overwhelmingly, psychologists recruit participants who are available
- They use *convenience* samples
 - The phrase is guilty of what Michael Billig calls “nouniness”: *convenient* would serve us here
 - We’ll stick with the customary expression, however
- Those who work at universities recruit students, or possibly other employees, or possibly persons who live nearby, unless no one from the population of interest dwells there
- So an old saying is still half true



HOW DOES CONVENIENCE SAMPLING AFFECT OUR STUDIES?

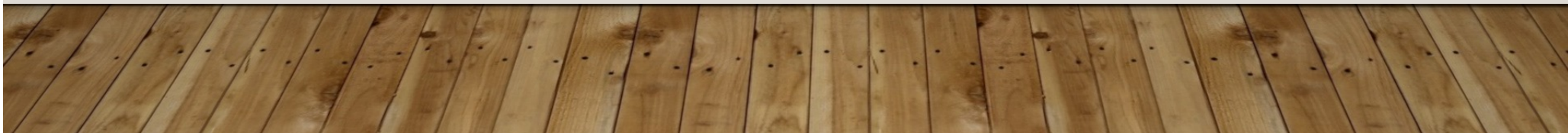
- Psychologists can be made to admit to using convenience samples
- It's the implications of using them that are essentially don't get talked about
- They need to be, because *none of our statistical procedures were developed for use with convenience samples*
- The difference between a convenience sample and a random sample is not like the difference between a normal distribution and one that is measurably leptokurtic

ARE CONVENIENCE SAMPLES OK?

- A reasonable guess: Psychologists act as though convenience samples are OK because everyone in the population is pretty much the same *with regard to the attributes being studied*
- Generalization will then be possible from a convenience sample to the population (what inferential statistics will be left to contribute is another matter)

OF COURSE, WE RARELY KNOW WHEN OUR ASSUMPTIONS ARE CORRECT

- Suppose you collect data on New Zealand college students who have caught a respiratory virus of unmentionable origin
- Will your results generalize to preschoolers or middle-aged or old people (in New Zealand or elsewhere)?
- If you've acted as though people at every age will be equally likely to catch the virus, to be hospitalized on account of it, or to die from it... your assumption is spectacularly wrong

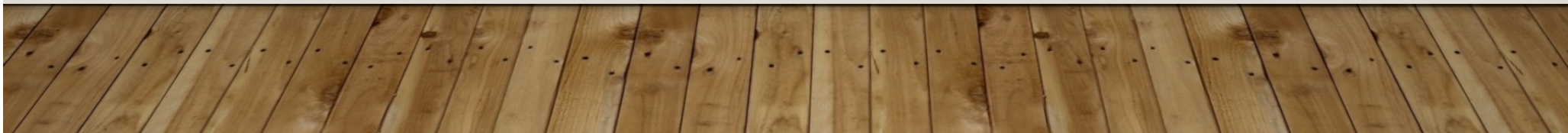


THE STATISTICAL CASE FOR REPLICATION

- If everyone quits QRPing tomorrow, empirical research in psychology will still be in trouble
- Because we regularly apply statistics that assume random sampling to convenience samples, we (again) have no realistic option but to *replicate, replicate, replicate*

WHILE CONDUCTING NEEDED STUDIES AND DOING THEM BETTER, LET'S IMPROVE OUR COMMUNICATION

- Billig, Michael. 2013. *Learn to write badly: How to succeed in the social sciences*. Cambridge: Cambridge University Press.
- “My charge is not that academic social scientists happen on occasion to use technical terminology in less than desirably clear ways, as if sometimes they unguardedly let standards slip. The charge is stronger: academic social scientists have created terminology that is inherently ambiguous.” (p. 84)



DE-AGENTIALIZATION
AGENTIALIZATION
MEDIATIZATION
NOMINALIZATION
DISCINOPTIC LAMINATION
MASSIFICATION
FUNCTIONAL METAFUNCTION
RELATIONAL CATEGORIZATION DEVICE
GOVERNMENTALITY
AUTOMATICITY
RE-ETHNICIFICATION
DESCRIPTIVIZATION
TRANSITION RELEVANCE PLACE
HABITUALIZATION
PASSIVIZATION

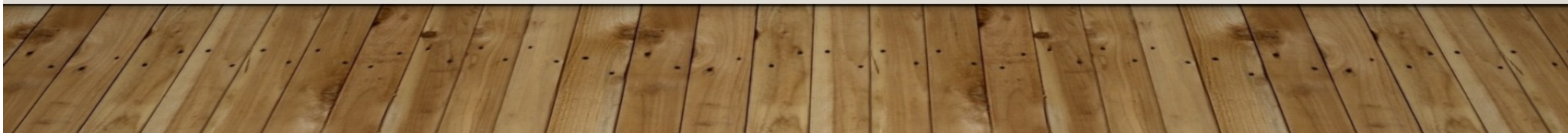
Learn to Write Badly

How to Succeed
in the Social Sciences

Michael Billig

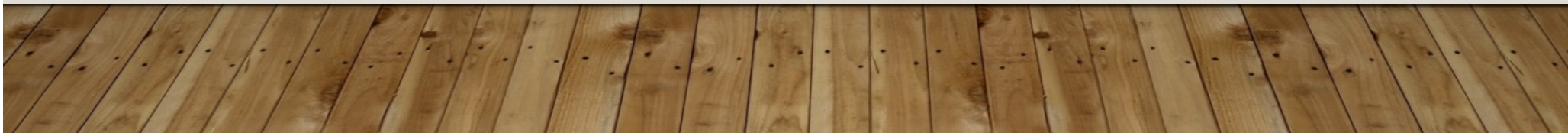
THE CORE PROBLEM: NOUNINESS

- “the big concepts which many social scientists are using — the ifications and the izations — are poorly equipped for describing what people do. By rolling out the big nouns, social scientists can avoid describing people and their actions. They can then write in highly unpopulated ways, creating fictional worlds in which their theoretical things, rather than actual people, appear as the major actors. [...]”
- **‘The paradox is that both bureaucrats and natural scientists use heavily nouny styles,** often because they can avoid specifying who is doing what. In my view **it has been disastrous for social scientists to follow them.’**
- —*Learn to Write Badly*, p. 7



PINKER: PSYCHOLOGISTS AREN'T WRITING BADLY ON PURPOSE

- Pinker, Steven. 2014. *The sense of style: The thinking person's guide to writing in the 21st century*. New York: Viking.
- “The curse of knowledge is the single best explanation I know of why good people write bad prose. It simply doesn't occur to the writer that her readers don't know what she knows—that they haven't mastered the patois of her guild, can't divine the missing steps that seem too obvious to mention, have no way to visualize a scene that to her is as clear as day. And so she doesn't bother to explain the jargon, or spell out the logic, or supply the necessary detail.” (p. 61)



BILLIG: YES, THEY ARE

- “Over the years, social psychologists have developed conventional ways of doing research which aspiring doctoral students must acquire. Successful careers will be the reward for following these procedures assiduously and for obeying the productive logic of the world of variables. But there is one thing that success will not bring: that you will actually know what is going on the experiments that you read about, or, indeed, in the experiments that you conduct. In fact, *your expertise will teach you how not to know what is going on.*”
- —*Learn to Write Badly*, pp. 190-191, italics added

