DOCUMENT SUMMARY This review from *Frontiers in Psychology* addresses a fundamental logical problem in psychological research, termed the "**ergodic fallacy**." The authors argue that the common practice of using group averages (e.g., means) to make generalizations about individuals is invalid because human beings are non-ergodic systems. This mismatch between sample-level statistics and individual-level reality threatens the validity of many psychological theories and their application. The paper proposes a straightforward method of "**pervasiveness analysis**" as a solution, which involves counting the proportion of individuals in a sample who actually exhibit an effect, thereby providing more valid and useful data for making claims about behavior.

FILENAME speelman_2020_research_review_research_methods_ergodic_fallacy_pervasiveness.md

METADATA Category: RESEARCH Type: review Relevance: Core Update Frequency: Static Tags: #research_methods #ergodic_fallacy #validity #replication_crisis #quantitative_psychology #statistics #individual_differences #pervasiveness Related Docs: Documents discussing research methodology, statistics, the replication crisis, and the generalizability of research findings to individual clients. This paper provides a critical lens for evaluating other research. Supersedes: N/A

FORMATTED CONTENT

INTRODUCTION

Much of psychological research suffers from a logical problem called the

ergodic fallacy. The methods used to evaluate research examine group-level phenomena (like means) but provide little information about the experiences of individual people. This is a fundamental problem with the validity of generalizing from samples to individuals. This paper outlines the

ergodic fallacy, illustrates the problem with examples, and proposes an analysis of "**pervasiveness**" as a more valid way to understand psychological effects.

PSYCHOLOGY'S ERGODICITY PROBLEM: INDIVIDUALS ARE NOT GROUPS

Most psychological research involves studying groups of people to learn about a population. However, researchers often extrapolate from the population to draw conclusions about the individuals who constitute it. We measure collections of individuals to theorize about single individuals, which is problematic.

Using large samples and calculating averages is a standard method to minimize error and increase statistical power. However, the conclusions drawn from these sample-level statistics must also remain at the sample level. Mean-based statistics only allow generalizations to the means of other samples, not to individuals.

Generalizing from a group to its individual members is only valid if the system is "

ergodic". The ergodic theorem states this is only possible when two criteria are met:

1. All members of the sample are essentially interchangeable.
2. The members and the sample's average behavior do not change over time.

    Human beings are quintessential examples of non-ergodic systems. No two people are interchangeable, and people constantly change and learn over time. When psychologists use sample averages to describe individual psychological function, they commit the

    ergodic fallacy by treating a non-ergodic system as if it were ergodic. In a non-ergodic system, there is no reliable connection between sample-level statistics and the individuals in the sample. We can generate means from raw data, but we cannot generate those raw data from their means.

A recent study by Fisher et al. cite_start confirmed this, finding that individual participants were two to four times more variable than the samples they belonged to.

SCENARIO 1: WHAT CAN AN EXPERIMENT TELL US ABOUT REAL-WORLD BEHAVIOR?

Imagine an experiment comparing a new study technique to a traditional one. Participants are randomly assigned to two groups, take a pre-test on a new topic, use their assigned study technique for 30 minutes, and then take a post-test. If a statistical test (like a t-test or ANOVA) shows the new technique group improved significantly more than the traditional group, a typical conclusion would be: "The new study technique led to better learning".

This conclusion is often extrapolated to suggest that

any individual would be better off using the new technique. This is an example of the

ergodic fallacy. The experiment only provides information about samples, not individuals. Within the group that showed a significant average improvement, it's possible that some participants' performance was unaffected or even worsened. Standard analysis of mean differences suppresses this individual variability. An educational psychologist applying this research to help a specific student has no way of knowing the probability that the new technique will work for that individual.

This issue also calls into question the common assumption in cognitive psychology that we all share the same cognitive system. Models of cognitive processes are tested on the average performance of samples, not individuals. If individual performance were reported, it would often be clear that not everyone exhibits the same pattern as the sample average, which would count against a general, universal model of cognition.

SCENARIO 2: HOW PERVASIVE IS A SIGNIFICANT EFFECT?

Consider four hypothetical datasets (A, B, C, and D), each with 50 participants' improvement scores from a pre-test to a post-test.

- In
  Set A, all 50 participants show improvement.

- In
 Sets B, C, and D, the data is manipulated so that the *mean improvement is identical* to Set A, and a t-test for each set shows a statistically significant improvement.
- However, the individual data tells a different story. In
 Set C, some people's performance worsened, while others improved dramatically. In
 Set D, this pattern is even more extreme.

The point of this scenario is to demonstrate that, when we use sample-level statistics, an effect that is very large in some individuals can outweigh null or opposite effects in others, resulting in a significant overall effect.

A researcher looking at any one of these datasets would likely conclude they had found a "general effect," especially since common effect size measures would still be large. However, only Sets A and B would give us confidence in a universal effect. Standard effect size measures are summary statistics that reflect group characteristics and can be insensitive to dramatic differences among individuals. Presenting individual data is far more informative.

ANALYSIS OF PERVASIVENESS

To address the **ergodic fallacy**, researchers can perform an **analysis of pervasiveness**. This is not a new method, but a straightforward approach to better align data analysis with the conclusions researchers want to make. What is required is an indication of how many people in a sample were observed to behave in a specific way.

To conduct a pervasiveness analysis, three things are needed:

1. **Precise Definitions of Behavior:** A researcher must pre-define a clear, measurable criterion for what constitutes an "effect" at the individual level. This includes how large a difference must be to be considered meaningful.
2. **Count the Number of People:** Once the criterion is set, it is a simple matter to count how many individuals in the sample met it. This can be reported as a raw number or a proportion of the total sample.
3. **Define 'Majority':** A benchmark value should be pre-defined to determine if an effect is truly pervasive (e.g., 80% of the sample). This provides a degree of confidence that the observed phenomenon is characteristic of "most people".

Worked Example: Skin Lesion Detection Study

A re-analysis of an experiment on improving skill in detecting dangerous skin lesions (Speelman et al., 2010) illustrates this method. The original study used ANOVAs on group means and concluded that certain training conditions improved accuracy by 12-15%.

A

pervasiveness analysis was conducted on the original data.

- For reaction time (RT), it was found that "most" people (defined as ≥80%) in the effective training conditions improved their speed by at least 30%.
- For accuracy, however, the results were different. While the
 mean improvement was reported as 12-15%, the **pervasiveness analysis** showed that the greatest improvement exhibited by 80% of participants was only 2.5%.

This discrepancy demonstrates how reliance on means can be misleading in representing majority performance and can be inflated by outliers. The

pervasiveness analysis is less affected by such outliers and provides a clearer picture of the effect's consistency.

ADVANTAGES OF A PERVASIVENESS ANALYSIS

1. **Superior Measure of Effect Size:** A pervasiveness analysis indicates the size of the effect demonstrated by *most people* in a sample, unlike traditional effect size measures which are based on group averages.
2. **Results Closer to What We Want to Say:** It provides direct evidence to support claims that "most people behave in a certain way," bringing scientific conclusions closer to the intended meaning.
3. **Easily Combine Estimates:** Pervasiveness values from different studies (which are simple counts or proportions) can be easily combined in a meta-analysis to produce a more convincing overall pervasiveness estimate.
4. **Similar to Replicability:** The pervasiveness of an effect is a measure of its replicability; it represents the probability of observing the effect if one more person were tested.
5. **Indicates Homogenous Sub-Groups:** When an overall effect is not pervasive, a closer look at individuals may reveal sub-groups (e.g., based on a moderating variable) that behave in different, more consistent ways.

CONCLUSION

The

ergodic fallacy helps to explain the reproducibility crisis in psychology. If sample-level effects are just reflections of an idiosyncratic mix of individual behaviors, it is no surprise that they are difficult to replicate with a new sample of different individuals.

Researchers should worry if their experimental design and analysis methods assume that all participants share the same basic psychological mechanisms and if those methods mask individual variability. If so, their conclusions are not valid at the individual level.

We have presented a simple method for determining the

pervasiveness of psychological effects. While systemic change is needed to "fireproof" the discipline against the

ergodic fallacy (e.g., changes to education and publishing norms), **pervasiveness analysis** is a "fire-fighting" measure that researchers can apply immediately to existing datasets. This approach can unlock untapped value from data we already have and help psychology produce results that are more valid, reliable, and applicable to the individuals we aim to understand.