***In Search of a Method***

# Chapter 1

**CHAPTER 1 Contents:**

1.1 What is Artificial Psychology?

[1.2. In Search of a Method 13](#_Toc125221899)

[1.3. From p-value to p-war 14](#_Toc125221900)

[1.3.1. P-value as Evidence of Prove or Disprove of a Null Hypothesis 15](#_Toc125221901)

[1.3.2. Reverse Interpretation of p-value 16](#_Toc125221902)

Abstract: We present the assumptions and philosophy underlying artificial psychology (AP) and motivate the need for a set of models which can incorporate information from complex mental systems using ideas such as fuzziness of a system and unsupervised algorithms which use artificial intelligence. We discuss the need for a multiplicity of modelling approaches to help us to understand the world. We mention issues involving hypothesis testing, in particular we introduce and define the p-value and highlight its shortcomings including p-hacking where data is manipulated so that it yields statistically significant results with an associated bias towards publishing studies which have p-values below a certain threshold. We mention widespread misunderstandings in the interpretation of the p-value and associated dangers such as giving the impression that the world is ‘black and white’ and motivate the need for complementary more nuanced approaches to testing statistical hypotheses which can overcome these deficiencies.

Keywords: Statistical inference; Artificial psychology; P-values; Hypothesis testing; Uncertainty

**1.1 What is Artificial Psychology?**

Artificial Psychology (AP) is a highly multidisciplinary field of study in psychology. AP tries to solve problems which occur when psychologists do research and need a robust analysis method. Conventional statistical approaches have deep rooted limitations. These approaches are excellent on paper but often fail to model the real world. Mind researchers have been trying to overcome this by simplifying the models being studied. This stance has not received much practical attention recently.

Promoting and improving artificial intelligence helps mind researchers to find a holistic model of mental models. This development achieves this goal by using multiple perspectives and multiple data sets together with interactive, and realistic models. This comprehensive, holistic, and interactive view may lead to a new research line in the near future. AP can open up a new horizon for mind researchers from clinical to theoretical psychologists to find a more realistic model. This horizon is rooted in a multidisciplinary approach updating our view along with development of the related sciences leading to the finding of new results even from old datasets and models. AP has some assumptions. Satisfying these assumptions helps find a more precise and deeper way of modelling for artificial psychologists.

The assumptions of AP are discussed here. Firstly, that the mind is filled with uncertainty. The uncertainty is the cost we are paying for living in the real world. We are usually trying to proceed through this uncertainty by considering the most certain fact as a truth (Shukla, Tiwari & Kala, 2020). It is important to note that uncertainty not only occurs in nature but also in almost all man-made systems. Secondly, we assume that mind is continuous. In other words, we assume a continuous consciousness in which the brain acts holistically and outputs behaviors discretely (Huette,2012) therefore, there is not a sharp dividing line between emotion and cognition. The brain is a grey matter which constructs mental systems which are not separated by solid lines. These ambiguous areas are the ones mind researchers are trying to handle by the use of statistical models. The third assumption is that mind is a complex system, human mentality is made up of complicated systems. Even the simplest system and phenomena are complex. This complexity can be captured and interpreted by a dynamic model. The fourth assumption is that there is always a proxy between mind and data. It is not possible to study mental activities directly. Brain data needs to connect to some psychological constructs and behaviors. Therefore, we need to use multiple sources of data in a single model at the same time. Conventional statistical techniques use rigorous mathematical models. These models require comprehensive and complete data for analysis and prediction. In the real world, we are facing big, imperfectly measured data as well as nonlinear relationships in complex systems. The fifth assumption is that brain data is highly dimensional data. This implies the dataset has many features even in small sample sizes. This problem commonly occurs in psychological research, especially in clinical, cognitive psychology and neuroscience where we need to deal with P>n.

In summary, psychologists need new analysis models which can help them to model complex mental systems. Artificial psychology uses intelligent models which satisfy these assumptions (Figure 1).

1. Artificial Psychology as a multidisciplinary field

One technique used in applied computing is to emulate the strategies involved in the intelligent systems or models for problem solving. Intelligent models are related to the human way of thinking and interpretation. These models use fuzzy logic, artificial intelligence and genetic algorithms both individually or together.

## 1.2 In Search of a Method

Science aims to clarify concepts systematically, and its core is replicability. The ambient world is full of concepts, some of which are determinable or almost determinable, such as country borders and per capita income. Others, however, that are more prominent and studied in the social and behavioral sciences are different, such as depression, suffering, grief, love, and selective attention.

There is an infinite set of these concepts and they are very vague, scattered and elusive and assume highly diverse intertwined forms. The models designed by researchers to acquire an approximate understanding of these concepts represent the efforts of science to clarify them. Models are an approximate simplified understanding of reality and do not deal directly with reality. Models are always an approximation of reality outside of them. These models test theories and hypotheses about different concepts.

If psychological reality is vague, and elusive, how should it be examined then?

Such a reality cannot be understood well by a single method. Therefore, research methods must develop like any other science, and there is no harm in employing a multitude of methods and, sometimes, methods of other disciplines for more accurate proximity to concepts.

Perhaps the principles and assumptions of many of the existing methods and future ones will probably require revision, even the methods introduced in this book. The dominance and hegemony of a particular method is remarkably alarming and dangerous.

It should be borne in mind that methods are not the only means of achieving reality but they are the containers of reality, and the research findings take shape from them. Furthermore, these findings from the method occasionally become so extreme that they annoy some researchers.

It is noteworthy that the present book does not attempt to discard, ignore, or devalue past or existing methods but to overcome the fear of going beyond them. The research psychologists' shared fear might result in innovation. The aim is to find a way in which elusive concepts could be understood more clearly by researchers.

We know that wrong answers are more harmful than random answers because wrong, non-random answers mislead science systematically and significantly.

From Sigmund Freud's birth in 1856 until the writing of this book (2022), the earth has rotated 60,590 times, and research methods need to rotate as well. We know that psychology did not begin from Freud's birth, but Freud is cited since he has been called one of the scientific revolutionaries.

## 1.3 From p-value to p-war

Statistical significance plays a significant role in scientific research by linking data to hypothesis testing from the late mid-twentieth century (Gigerenzer et al., 1990). Currently, the most commonly used statistical measures in scientific studies, despite much criticism of its use, is the p-value (Lyu et al., 2020). In spite of the widespread use of the p-value in psychological research, various studies show that most researchers and students misinterpret p-values. This misinterpretation is rooted in p-value misuse, including statistical significance hypothesis (Ziliak & McCloskey, 2008) and p-hacking. These are the main reasons for the confidence crisis and reproducibility crisis in psychology research. The effect size and confidence intervals (CIS) can be regarded as alternatives to the p-value. Although the CIS can be used to improve statistical interpretation and inference, considering it an indicator of the effect size variations, its concept has also been misunderstood by researchers and its users (Lyu et al., 2020; Harrison et al., 2020; Greenland et al., [2016](https://journals.sagepub.com/reader/content/182ee65b514/10.1017/prp.2019.28/format/epub/EPUB/xhtml/index.xhtml#bibr22-prp-2019-28); Lyu et al., [2018](https://journals.sagepub.com/reader/content/182ee65b514/10.1017/prp.2019.28/format/epub/EPUB/xhtml/index.xhtml#bibr32-prp-2019-28); Morey, Hoekstra, Rouder, & Wagenmakers, [2016](https://journals.sagepub.com/reader/content/182ee65b514/10.1017/prp.2019.28/format/epub/EPUB/xhtml/index.xhtml#bibr38-prp-2019-28); Cumming,2013).

The study carried out by Lyu et al. (2020) on 1479 researchers and students in various fields in China revealed difficulty interpreting the p-value and CIS correctly, regardless of their academic degree and career stages. 89% of them made at least one error in the p-value interpretation, and 93% made at least one error in the CIS interpretation. The level of misinterpretation in the significant and non-significant p-values and whether the CIS included zero or not was increased. Moreover, it is noteworthy that respondents were generally confident in their (incorrect) judgments.

These results indicate that researchers have misunderstood these crucial indicators of inferential statistics. This misunderstanding causes researchers to misinterpret, using classical statistics-based methods (assuming we are pleased with the p-value!), and these interpretations flow from different streams into the sea of psychological research findings (Harrison et al., 2020).

These interpretations include the following:

### 1.3.1. P-value as Evidence to Confirm or Unconfirm a Null Hypothesis

This issue is what is referred to as the illusion of certainty in Gigerenzer's research (2004, 2018), as it may provoke a crisis of confidence in the psychological research findings by encouraging researchers to reach a p-value ≤ 0.05 as evidence of the existence of an "effect." One of the primary sources of this crisis is publication bias. This bias results from the fact that scientific journals welcome statistically significant results (p-value≤0.05).

Chang et al. (2019) state that a p-value is the probability of obtaining an effect at least as extreme as the effect in the sample data, assuming the truth of the null hypothesis. Considering the p-value in classical statistics, despite its low statistical power in both single studies and meta-analyses, may result in distrust in the actual results of psychological research. Given the statistical power in the published studies, the frequency of statistical significance in those studies is suspiciously high (Francis, 2014; Francis, Schimmack, et al., 2012).

Schmidt and Oh (2016) asserted that 90% of the research reports were significant, while the average total power was 0.4. What they found is strong evidence that these studies are questionable. Test power did not increase in questionable research practices until 1962, when Cohen emphasized the "low power" issue. What was the reason for this emphasis?

We know that in classical statistics, the p-value is a function of the effect size and the sample size, and the sample size seriously affects the power increase:

$$\left(p-value\right)=f\left(effect size, sample size\right)$$

**(1.1)**

Low power leads to non-significant results. Non-significant results are as important as significant ones, but their non-significance made it difficult for researchers to publish their papers. This publication bias is called the file drawer problem with studies less likely to reject the null hypothesis ending up unpublished in a file drawer. Therefore, researchers have earnestly strived to increase the power of their research. Maxwell (2004) has elaborated on this issue extensively. However, it has not always been possible to obtain a sufficient sample size, so maybe this is why researchers conducted questionable research. In other words, they conducted questionable research to obtain significant results to avoid the abundance of non-significant results due to low statistical power (Schmidt & Hunter, 2015; Barker et al., 2012, Harrison et al, 2020).

### 1.3.2 Reverse Interpretation of the p-value

Another consequence of misinterpretation of the p-value, as Lyu, (2020) states, is replication illusion. Many researchers avoid Bayesian thinking because of classical p-value-based statistics, the thinking that is the basis of classical inference. The reverse interpretation of the p-value is to consider 1–p-value as the probability of successful replication of the result.

Despite these problems, the potential consequences of the lack of statistical thinking and ritual use of p-values have rarely been mentioned in the psychological research results, except in recent years (Lyu, (2020). The study of Farahani, Azadfallah, and Roshan (2021) on a sample of 100 postgraduate and Ph.D. psychology students in Iran indicated that 95.7% of them make mistakes about the illusion of certainty and the replication illusion in the interpretation of the p-value.

The p-value is not well understood, and most researchers speak about it with a wrong mindset and perception. A p-value demonstrates the likelihood that the researcher's data will occur under the null hypothesis. This is obtained by calculating the likelihood of test statistics gained from the researcher's data (Indrayan,2019).

It should be noted that the p-value is a ratio and a percentage. The p-value is the probability of a test statistic at least as big as the test statistic obtained from the data, assuming that the null hypothesis is correct.

Harison and et al (2020) summarized the shortcomings of using NHST (p-values) as follows:

1. Use of p-values without regarding the effect sizes and confidence intervals is not informative

2.The potential for use of ‘p hacking’ by manipulating data and analyses deliberately to reduce p-values

3.Simplistic dichotomous interpretations of p-values as either significant or nonsignificant

4.Incorrect interpretation of p > 0.05 as no effect

6.Misinterpreting statistical significance and taking it as clinical or practical significance

7.Committing multiplicity by performing multiple statistical tests without adjusting the criterion p-value.

One way to improve p-value interpretation is to use clinical interpretation, practical interpretation, or practical significance. Apart from statistical significance, the effect size should be used for practical interpretation. Another point is that reporting inconclusive findings and null findings in articles is not harmful but valuable and strengthens the scope of scientific theories, but it should be borne in mind that what was said at the beginning of this chapter about the world not being black and white encourages researchers to choose another way to have accurate yet close-to-reality findings.

To design a different research model, a new conceptual framework is required with different measures, which will be discussed in detail in the second chapter.