Introduction

Some empirical researchers consider objectivity to be an essential part of scientific methods that yield bias-free results (Dettweiler, 2019). A possible explanation for the popularity of statistics is its claim of making objective inferences (Daston & Galison, 2021). To claim objective inference is to claim a view from nowhere, i.e. free from personal bias and values.

Statistics is a mathematical and conceptual discipline that examines the association between collected data and hypotheses about the data (Romeijn, 2017). One key goal of statistics is to allow for inferences from the data that help us explain the data (i.e. statistical reasoning) (Godfrey-Smith, 2009). Null hypothesis significance testing (NHST) is presumed to yield objective results (by some, but not all researchers) (Morawski, 2021) and is widely used in empirical psychology (Cristea & Ioannidis, 2018).

NHST has been considered to be the sine qua non of scientific research (Gigerenzer, 1993), and aims to test the null hypothesis: the probability of the observed data occurring, assuming that a specific hypothesis is true (Krueger & Heck, 2017). However, researchers' use of NHST has been criticized for the misapplication and misinterpretation of P-values (Cohen, 1994; Krueger & Heck, 2017) and to make objective inferences, claiming that the results either "prove" or "disprove" an hypothesis (Amrhein et al., 2019). In other words, that some researchers apply NHST ritualistically and mechanically, focusing too much on P-values without comprehending their meaning (Krueger & Heck, 2017). Gigerenzer (2018) coined this the "null ritual", i.e. researchers' faith in statistical methods and subsequent inferences that appears to eliminate judgments. Because of the ritualistic use of NHST, researchers argue that it is a result of epistemological unclarity (Hanfstingl, 2019; Meehl, 1997), and wrongful interpretation of P-values and their significance (Cohen, 1994; Gelman & Loken, 2013).

NHST is suggested to have been incentivized in empirical psychology when scientific progress is measured by the accumulation of significant effects (Simmons et al., 2011). However, Proulx and Morey (2021) suggest that the replication crisis in psychological sciences focuses too much on statistical methodology at the expense of theory-development. One particularly relevant point is the theory-ladenness of scientific observations and instruments, i.e. researchers' theoretical presuppositions guide what they look for (Fjelland,

1991). This includes statistical methodology and underscores how scientific concepts derive from theories. Hence, Proulx and Morey (2021) suggest that psychological sciences should bring theory-development to the forefront and acknowledge that statistics is a way of reducing data to manageable forms. It should be mentioned, nonetheless, that a replication crisis may be a natural part of the scientific enterprise (Shrout & Rodgers, 2018) and that other fields, such as physics, chemistry and medicine, also face reproducibility challenges (Baker, 2016).

A less debated cause of the replication crisis in psychology concerns the relationship between philosophical preconditions and NHST (Morawski, 2019). In their everyday scientific inquiries, scientists are guided by what they perceive the world to be (ontology) and what they think they can know about it (epistemology) (Andersen et al., 2019). These are known as philosophical preconditions to which scientists implicitly adhere. Thus, philosophical preconditions guide what we see and look for and what we judge as possible to know in science.

This article analyses whether the ritualistic use of NHST in empirical psychology may be influenced by the philosophical precondition of an ontology of numbers. Moreover, it suggests that an ontology of numbers is closely related to the epistemological precondition of scientific objectivity, which could lead to the ritualistic use of NHST, thus contributing to the replication crisis. A brief overview of the main points to be developed within the further sections follows:

The “null ritual” may contribute to the replication crisis in empirical psychology. The “null ritual” occurs when researchers misuse or misapply null hypotheses significance testing. Misuse or misapplication leads to an over-production of false positive non-replicable results. This causes replication problems. One possibility for the null ritual is the researchers’ lack of awareness of philosophical preconditions. Especially the ontology of numbers which is fundamental to statistics. Furthermore, unawareness of philosophical preconditions may put researchers at risk of ontological reductionism, i.e. the whole of reality stemming from a minimal number of parts. I discuss this in relation to the epistemic virtue called “scientific objectivity”. It is possible that null significance hypotheses testing includes many of the characteristics of scientific objectivity that may mislead researchers to think that this method is objective. I propose that this risk may be mitigated through the researcher’s

acknowledgement of a different ontology (i.e. their "lifeworld") and how this ontology constitutes their scientific world. This may contribute to methodological reductionism rather than ontological reductionism. I suggest that focus on history and philosophy of science, both to students and faculty employees, might be part of the solution.

The replication crisis and null hypothesis significance testing

The replication crisis refers to a number of challenges, such as the apparent absence of replication studies in psychological research (Makel et al., 2012), evidence of publication bias (Fanelli, 2010), or the failure of sizeable systematic replication studies to reproduce published results (Begley & Ellis, 2012). Recently, others have mentioned that scientific publications are not transparent enough and incomplete in their reporting of the methods, findings and analyses (Nuijten et al., 2016), and too much questionable research practice (Fraser et al., 2018) as additional reasons for the replication crisis. Shrout and Rodgers (2018) suggest that good scientific practice may improve replicability in psychology, such as researchers adopting open science conventions of preregistration and more sophisticated statistical analyses. However, open science practice has been met with critique, if it leads to excluding specific studies due to lack of transparency, which is suggested to impede decision-making processes (Berg et al., 2018).

Szucs and Ioannidis (2017) claim there is an over-production of false positive non-replicable results. One reason for this over-production is that researchers are mainly educated in NHST and tend to misunderstand and misuse NHST. They argue that one worrisome belief is that some researchers believe that NHST (or any statistical tool) can disprove or prove a hypothesis once and for all. Szucs and Ioannidis (2017) argue that researchers' misunderstanding of the probability of producing false positive findings leads to overconfidence in research findings and contributes to the replication crisis in psychology. This resembles the null ritual in the sense that some researchers use NHST incorrectly, believing NHST can prove or disprove a hypothesis. Ritualistic use assumes that NHST contains a set of actions that may be routinely applied and that will help the researcher to decisively prove or disprove a hypothesis. Thus, it seems that some researchers use NHST without good judgment.

Since NHST is widely used in empirical psychology (Cristea & Ioannidis, 2018) and commonly stated as one of the primary contributors to the replication crisis in empirical psychology (Szucs & Ioannidis, 2017), this paper focuses on NHST's relation to the replication crisis in psychology. Specifically, by focusing on the philosophical precondition of an ontology of numbers and its relationship to scientific objectivity, and how this relationship may increase scientists' belief(s) in the certainty of their results. This could make some of them more at risk to conduct the null ritual. Nevertheless, it should be mentioned that the concepts of "replication", "reproducibility" and "repeatability" are related and often used synonymously, although they may be distinct from one another. This paper uses the concept "replication", understood as the possibility of conducting the study again and the possibility of the replicated study to produce (satisfactorily) similar results.

A central aim of the quantitative methodology is "explaining phenomena by collecting numerical data that are analyzed using mathematically based methods (in particular statistics)" (Muijs, 2004, p. 1). NHST represents such a mathematically based method. A prerequisite for using mathematical methods, such as NHST, are numerical data (Muijs, 2004). If one wishes to analyze qualitative aspects they must be transformed into numbers, e.g. categorical data is numerical representations of physical traits such as gender or hair color. The null hypothesis states that the study's hypothesis is wrong and that there is no effect (Field, 2013) and is the opposite of the experimental hypothesis. The hypothesis is judged as either significant or non-significant using a P-value. If the findings have a P-value ≤ 0.05, the findings are said to be significant. Cohen (1994, p. 998) argues that NHST states that "if the null hypothesis is correct, then these data are highly unlikely. These data have occurred. Therefore, the null hypothesis is highly unlikely". However, Cohen (1994) argues this is inaccurate due to the incorporation of probability and wrongful application of deductive logic, i.e. an incongruent application of a logically consistent *modus tollens* (Perezgonzalez, 2017).

Deductive *modus tollens* statements have this structure: if A then B. Not B therefore not A (Cohen, 1994). However, probabilistic *modus tollens* statements are suggested to be an incorrect form of inference within frequentism (Sober, 2008). One cannot assert that since a hypothesis states that an observation is highly unlikely that the hypothesis is improbable as NHST does (Sober, 2002). However, using Bayesian subjective probability theory, rather than frequentist theory as previously mentioned, it is possible to evaluate inferences from

uncertain premises (Evans et al., 2015). Whereas probabilistic *modus tollens* within a frequentist paradigm is too subjective and therefore often disregarded, in a Bayesian framework, it makes sense (Sober, 2008). In conclusion, an unlikely event does not yield evidence for, or justify, inferring that the particular event is improbable.

NHST and a significance level do not tell us the truth of a hypothesis, but the relative frequency of type 1 errors in the long run (Gigerenzer, 1993). Goodman (2008) contends that P-values less than 0.05 does not reflect scientific proof of a particular observation, but that it warrants more experimentation. Nevertheless, Cohen (1994) claims that some misperceive NHST to state the probability of the null hypothesis being true given the data. Cohen (1994) emphasizes that these statements are not the same. The latter may be interpreted as stating that a P-value of 0.05 means that the probability of the null hypothesis being false is 95%. This is called the *inverse probability fallacy*, which is the wrongful conviction that the probability of the data given the null hypothesis is equivalent to the probability of the null hypothesis given the data (Kalinowski et al., 2008).

The probability derived from NHST concerns hypothetical frequencies of data patterns given a particular statistical model and not hypotheses (Greenland et al., 2016). Focusing too much on NHST may divert our attention away from other statistical assumptions equally pertinent to interpreting the result's validity. Greenland et al. (2016) argue that the P-value provides information about the compatibility between observed data and our predictions, given that our statistical model was correct. Thus concluding about findings based only on P-values overlooks that other assumptions, such as philosophical and theoretical ones, influence the findings and P-values.

Reducing human phenomena through quantification

A common trait of science is reducing or simplifying of complex phenomena into more manageable parts. For example, NHST reduces complexity when converting human phenomena into numbers using quantitative methods. Experimental conditions, i.e. idealized settings and variable control, are used to increase the certainty of scientific results (Fjelland, 2002). This is one way of controlling the variables under investigation to make the

specific phenomenon more comprehensible and interpretable. However, when something is reduced, it is brought down in degree, amount or extent. Consequently, in a sense, reducing something means to remove or lose something else. However, to reduce or simplify something may or may not be a problem in science. This depends on what is meant by to "reduce" or to "simplify".

When discussing reduction in NHST, it is possible to distinguish between ontological and methodological reductionism. Ontological reductionism is the claim that the whole of reality stems from a minimal number of parts. To claim that a person's experience of joy or sorrow is "no more than" or "nothing but" chemical processes between nerve cells is an example of ontological reductionism (Fjelland, 2020). In relation to NHST in empirical psychology, an ontological reductionist would presume that reducing the phenomenon to numbers by measurements and then using statistics to analyze the numbers brings the phenomenon back to a more basic level (e.g. fundamental laws of human behavior).

Methodological reductionism, on the other hand, denotes the use of scientific methods in order to explain a phenomenon in smaller entities. A methodological reductionist would probably use NHST as a means to explain a phenomenon "A" by reducing it to some parts (B, C, D) without claiming that they represent a more fundamental level of "A". Hence, ontological reductionism claims to bring back the phenomenon to its more basic form, while methodological reductionism claims to explain a part of the (whole) phenomenon.

Both are related to scientific objectivity, but different aspects of objectivity as well. Ontological reductionism may be related to bias-free judgments and a reality independent of human observers, while methodological reductionism may be associated with standardized procedures for collecting data and resorting to quantification. Unfortunately, if scientists are unaware of them, reductionism may be dangerous for science. For example, scientists may presume that scientific methods yield objective representations of reality, when they actually yield perspectival representations. Moreover, they may think that mental phenomena are "no more than" chemical processes between nerve cells, overlooking that humans are bodily and social beings who live in a material and interpersonal world (Fjelland, 2020).

Statistical methods provide a way to study human activity by reducing complex behaviors

into numbers (methodological reductionism). Statistics may give us knowledge about a given phenomenon, but do not explain the entire phenomenon. The premise, or rather the philosophical precondition that human activity may be studied statistically, is primary to the assumptions of a statistical model. If this premise goes unnoticed, it may lead to ontological reductionism. In this regard, there is a risk of reducing complex human behavior to a level of analytic abstraction that overlooks the qualitative and practical aspects of human life (Canguilhem, 2000). Moreover, failure to be aware of philosophical preconditions may overlook the fact that scientific objectivity is an epistemic *virtue,* i.e. it relies upon distinct normative codes of scientific conduct for investigating nature.

The mathematical reduction in NHST may result in two challenges for empirical psychology. First, everyday human phenomena are transformed into theoretical and empirical abstractions separated from everyday human life. Thus there is a gap between the ideal world of science and the real world. Second, which is related to the first, the mathematical reduction will always involve a degree of uncertainty that presumably makes it difficult to reproduce the same study and produce the same or similar results (i.e. replication).

This degree of uncertainty in mathematical reduction concerns the gap between phenomenon and numerical representation. What we want to measure is not "pre-given". It rests upon researchers' judgments. Therefore, uncertainty arises related to how well the numerical representation reflects or represents the phenomenon. It should be mentioned that statistics was developed to make "lower sciences" such as psychology more exact (Fjelland, 2010). To use statistics to be more exact rests on the fact that statistics simplify and idealize human behavior. It reduces human behavior through mathematical theory.

However, these theories are not without fault. There is uncertainty related to NHST's random errors due to chance. Additionally, there are systematic errors, such as sampling bias, measurement error, and experimental error (Fjelland, 2022). In a complex system there is always a risk that our models of the world are incomplete. It is likely that psychological science is best understood as a complex system. Thus, using NHST to reduce psychology to fit a simpler system such as mathematical laws inevitably includes some degree of uncertainty. Both challenges are related to NHST's ontology of numbers. However, these challenges are not necessarily specific to empirical psychology, as most sciences involve abstractions and uncertainty.

The mathematization of human activity

At the heart of statistics lies the philosophical precondition that human activities may be transformed into numerical data that are analyzed using mathematically based methods. Thus statistics, and ultimately NHST, presume an ontology of numbers. According to Heidegger (1977, p. 289), the ontology of numbers presumes that real-world objects are determined by fundamental (mathematical) laws that may be found in those objects. This presumption gives rise to numerical measuring due to the claim that a universal uniform measure is an essential characteristic of objects (Heidegger, 1977, p. 293). Hence a universal blueprint is expected to be found in objects that may be explained using axiomatic propositions. The way statistics calculates and measures has implications for how scientists determine scientific objects. In this context, NHST presupposes numbers to be a universal constituent of the scientific objects and something that we can learn from them. However, it does not imply that researchers using NHST argue for ontological reductionism.

Sciences based on an ontology of numbers are often related to scientific objectivity. These sciences presume that the world may be described by using mathematical laws that objectively portray the world's real nature. The mathematization of nature and the mathematization of science are said to have originated in modern science (Koyré, 1943). For example, Galileo Galilei believed that the world was written in the language of mathematics (Galilei, 2016). He thought that quantitative methods could depict the real world without the influence of an observer. In recent times, others, such as Einstein and Hawking (Fjelland, 2002), as well as Crick and Harari (Fjelland, 2020), have made similar presumptions, claiming that the real world is reducible to more basic parts using mathematics (Anderson, 1972). They all were inclined towards ontological reductionism. However, as Anderson (1972, p. 393) points out, "*psychology is not applied biology*" (e.g. if we state that psychopathology is reducible to communication between neurotransmitters). At each level different considerations apply that are unique to that level.

NHST presupposes an ontology of numbers, i.e. that human activities may be transformed into numerical data that are analyzed using mathematically based methods. That scientists presuppose this is not controversial. However, it is controversial if scientists believe that

NHST gives them access to a universal blueprint, i.e. that real-world objects are determined by fundamental (mathematical) laws that may be found in those objects. In this sense, it seems to imply that NHST may give the scientists access to a "reality below" or the "real" laws that govern human behavior. Although we may use NHST to interpret psychological phenomena, it does not mean that human psychology is *mainly* governed by mathematical laws (or that higher level is transferable to a lower level). However, if scientists presume that NHST's numerical representation and calculation of human activity describe intrinsic aspects of humans they risk succumbing to ontological reductionism.

Mathematics is one hallmark of scientific objectivity. Other hallmarks are standardized procedures for registering data and resorting to quantification (Daston & Galison, 2021), which readily fits NHST, as well as statistical reasoning's embeddedness with rigor and certainty (Porter, 2020). Therefore, it may be that NHST's affinity to scientific objectivity, together with ontology of numbers, makes scientists more suggestible to conduct the null ritual.

In this respect, scientists who use NHST may neglect that how science describes the world is inherently interrelated with the everyday world of the scientist (Feyerabend, 1985) and that the proof for statistical propositions rests on experiences acquired through social interaction (Wittgenstein, 1969). We first and foremost perceive everyday objects before converting them into scientific objects (Moran, 2012).

Practical everyday life engagements

Heidegger (2010) presents a different ontology, arguing that our understanding of the world derives primordially from our practical everyday life engagements with the world as bodily and cultural beings. This is what Husserl (1970) coined the "lifeworld", which refers to the pre-given background of our lives and how we experience and interact with objects. Our "lifeworld" shapes the background that makes every form of human activity intelligible. To Heidegger (2010), our cultural upbringing and practical engagement with specific objects is the foundation for why we have a concept of specific objects.

In our everyday life engagements with things, the world is given to us in a subjectively relative way (Husserl, 1970). When Galilei mathematized nature, he presumed that the world was mathematical. However, he overlooked that to mathematize is a human act and a projection of already inherent presumptions. In this context, our "lifeworld" is a precondition for the scientific world. Humans interact with objects first and foremost, and not geometrical-ideal objects (Husserl, 1970). Galilei depicted an ideal world of science that is separated from our everyday world (Heidegger, 1977; Husserl, 1970). NHST is a statistical tool that scientists use to idealize psychology.

However, this is only a problem if it leads to ontological reductionism, i.e. if we mistake our everyday world for the ideal world of science. Such a mistake fails to notice that the ideal world of science is disconnected from our everyday practical situatedness in the world. However, as argued above, these worlds are inherently interconnected. When scientists use NHST, they should therefore recognize its ontological preconditions to illuminate that scientific objectivity is not bias-free and always includes uncertainty.

First and foremost, statistical models are simplified and idealized representations of reality. Any scientific experiment seeks to simplify the world by removing specific contaminating factors to increase certainty about which primary factors may be involved (Fjelland, 2002; Heidegger, 1977). Nevertheless, most scientists consider complete certainty or objectivity unattainable to science. One reason is that science is a representation of the world and not an identical depiction of it (Baudrillard, 2012). Although statistical models may increase knowledge about specific phenomena, they represent a simplified version of reality and thus may have low ecological validity (i.e. how well experimental results transfers to real-world settings). If this necessary simplification is overlooked, it may increase ignorance when results from statistical models are transferred back to reality (Fjelland, 2002), especially when relying solely on P-values. Transferring statistical results back to reality indubitably involves adding back contaminating factors that will increase complexity.

Increasing scientists' awareness of their lifeworld may positively impact empirical practice. Inevitably, the scientists' lifeworld is integrated with their philosophical preconditions and their scientific methodology. Knowledge of this integration may make it easier to notice their own premises for science. For instance, frequentists can infer from the group level to the individual. Although this is a valid conjecture from a frequentist position, it has been

associated with the "ecological fallacy", i.e. deciding what psychological treatment one person should receive based on what works best for a group.

If scientists are aware of their philosophical assumptions, I believe they may be more inclined to assess them in their ongoing research process. It will likely increase their awareness of how NHST, and other methods, are based on several assumptions that may include error and uncertainty. Thus, if scientists are aware of their philosophical preconditions, it may reduce misunderstanding and misapplication of scientific methods such as NHST. Consequently, this may reduce over-production of false positive results that contribute to the replication crisis.

How may awareness be increased?

Teaching and experience may be two possibilities. A course in history and philosophy of science to social science students, in Ph.D. programs and academic positions dedicated to philosophy of science in social sciences at universities. This could potentially increase knowledge about philosophical preconditions for doing science. For instance, what is the difference between modern and postmodern science? One difference is that modern sciences focused on scientific truths about the world, while postmodern science emphasizes the premises for scientific knowledge in a certain scientific discipline and the strengths and limits thereof (Lyotard, 1984). Researchers could also learn about how Galileo Galilei mathematized nature and distinguished between primary and secondary sensory qualities. As mentioned, Husserl criticized Galilei for overlooking the preconditions of science and the scientific method. Such topics are associated with measurement and mathematics in modern science (which is part of a scientific ideal that originated in the seventeenth century (Fjelland, 2022)). History of statistics should also be included.

Statistical methods, such as NHST, include value judgments and personal bias (e.g. thresholds for acceptance and rejection) (Reiss & Sprenger, 2020). The history of significance tests indicates, according to Ziliak and McCloskey (2009), that psychology adopted a simplified understanding of significance testing. This partly happened through the Publication Manual of the American Psychological Association and its overfocus on

narrow significance testing, while simultaneously downplaying other relevant information for statistical inference.

Social science students could learn about scientific values and norms. For instance, critically examine scientific objectivity and its desirability. Such investigations might delve into the limits of science, scientific knowledge, error and uncertainty. These topics scaffold with crucial researcher characteristics, such as humility, wisdom, and good judgment (Fjelland, 2022). Good judgment when using NHST may lead to reflections around measuring, and how measuring is related to what we want to know. Presumably, raising questions about NHST's foundation. This is favorable, since the null ritual may eliminate judgement (Gigerenzer, 2018). In addition, scientific experience (doing science in practice), working with experienced colleagues, participation in the scientific community and workplace, could potentially foster necessary aid to mitigate ritualistic use of NHST. Yet, this is probably only one part of the solution.

Conclusion

This paper advocates that the philosophical precondition of NHST of an ontology of numbers is vulnerable to ontological reductionism. The mathematizing of human activity may be prone to consider the real nature of human phenomena to be governed by mathematical laws. Furthermore, NHST has a strong affinity with the presumption of scientific objectivity facilitated by an ontology of numbers, but also the use of standardized procedures and quantification. Scientists who use NHST may therefore be particularly prone to the "null ritual" if awareness is lacking of its philosophical preconditions. However, mathematics is an essential feature of science. In this context, the paper argues in favor of methodological reductionism and ontological anti-reductionism, i.e. science yields perspectival and limited knowledge that is not reducible to a minimal number of parts. Thus statistical analyses may give us valuable knowledge about the average person.

The considerable ritualistic and mechanistic use of NHST (Cristea & Ioannidis, 2018) suggests that some researchers may be overly confident and certain about the validity and reliability of the P-values produced. Although this may indicate the epistemological unclarity

and misapplication of P-values by some researchers or even a consequence of publishing pressure (Gandevia, 2018), it also seems to neglect the implied philosophical preconditions and epistemic virtues (e.g. scientific objectivity). The "null ritual" suggests that some scientists rely too heavily on mathematical techniques and laws. This overlooks the fact that measurement procedures, quantification and statistical modeling are normative, i.e. value-laden. Although science presupposes that the world presents a minimum of a system (Kant, 2000) and NHST may be a useful way to study human activity, NHST contains limitations (Krueger & Heck, 2017).

Researchers should be aware of their lifeworld and philosophical preconditions of their practice. This will probably lead them to view NHST as a statistical tool that may help them to establish regularities, but not to disprove or prove hypotheses. This is in line with methodological reductionism. Awareness of our philosophical preconditions could also mitigate ontological reductionism. It will probably become clearer that ritualistic use implies that researchers consider there to be a "real level below" that is governed by mathematical laws that they can gain access to with NHST. This seems probable as some scientists consider NHST to disprove or prove hypotheses. Thus awareness of their lifeworld could make them more prone to consider an ontological *anti*-reductionist position (i.e. human behavior cannot be reduced to nothing but mathematical laws). In this context, they will likely be more attentive to how NHST cannot disprove or prove hypotheses and therefore be more aware of the possibility of producing non-replicable results. In this regard, scientists may be less likely to misunderstand and misuse NHST. Thus, also potentially moderate over-production of non-replicable results as mentioned by Szucs and Ioannidis (2017).

In conclusion, scientists, and scientific communities, must be aware of philosophical preconditions and implied epistemic virtues to counteract the ritualistic use of NHST. This may improve epistemological clarity and the application of P-values and facilitate a more thorough reporting of research, which in turn could increase the possibility of replication.

Declaration of interest statement

None.

References

Amrhein, V., Greenland, S., & McShane, B. (2019). Scientists rise up against statistical significance. *Nature*, *567*(7748), 305-307 https://doi.org/10.1038/d41586-019-00857-9

Andersen, F., Anjum, R. L., & Rocca, E. (2019). Philosophy of Biology: Philosophical bias is the one bias that science cannot avoid. *Elife*, *8*, e44929. https://doi.org/10.7554/eLife.44929

Anderson, P. W. (1972). More is different. *Science*, *177*(4047), 393-396.

Baker, M. (2016). 1,500 scientists lift the lid on reproducibility. *Nature*, *533*(7604), 452-454. https://doi.org/10.1038/533452a

Baudrillard, J. (2012). *Impossible Exchange*. Verso Trade.

Begley, C. G., & Ellis, L. M. (2012). Raise standards for preclinical cancer research. *Nature*, *483*(7391), 531–533. https://doi.org/10.1038/483531a

Berg, J., Campbell, P., Kiermer, V., Raikhel, N., & Sweet, D. (2018). Joint statement on EPA proposed rule and public availability of data. *Science*, *360*(6388), eaau0116. https://doi.org/doi:10.1126/science.aau0116

Canguilhem, G. (2000). *A Vital Rationalist: Selected Writings From Georges Canguilhem* (A. Goldhammer, Trans.; F. Delaporte, Ed.). Zone Books.

Cohen, J. (1994). The earth is round (p<. 05). *American psychologist*, *49*(12), 997-1003. https://doi.org/10.1037/0003-066X.49.12.997

Cristea, I. A., & Ioannidis, J. P. (2018). P values in display items are ubiquitous and almost invariably significant: A survey of top science journals. *PloS one*, *13*(5), e0197440.

https://doi.org/10.1371/journal.pone.0197440

Daston, L., & Galison, P. (2021). *Objectivity*. Zone Books.

Dettweiler, U. (2019). The rationality of science and the inevitability of defining prior beliefs in empirical research. *Frontiers in psychology*, *10*, 1866. https://doi.org/10.3389/fpsyg.2019.01866

Evans, J. S. B. T., Thompson, V. A., & Over, D. E. (2015). Uncertain deduction and conditional reasoning [Original Research]. *Frontiers in psychology*, *6*(398). https://doi.org/10.3389/fpsyg.2015.00398

Fanelli, D. (2010). Do pressures to publish increase scientists' bias? An empirical support from US States Data. *PloS one*, *5*(4), e10271. https://doi.org/10.1371/journal.pone.0010271

Feyerabend, P. K. (1985). *Realism, Rationalism and Scientific Method: Volume 1: Philosophical papers*. Cambridge University Press.

Field, A. (2013). *Discovering Statistics Using IBM SPSS Statistics*. Sage Publications Inc.

Fjelland, R. (1991). The theory-ladenness of observations, the role of scientific instruments, and the Kantian a priori. *International studies in the philosophy of science*, *5*(3), 269-280. https://doi.org/10.1080/02698599108573399

Fjelland, R. (2002). Facing the problem of uncertainty. *Journal of Agricultural and Environmental Ethics*, *15*(2), 155-169. https://doi.org/10.1023/A:1015001405816

Fjelland, R. (2010). The Problem of Scientific Uncertainty. *SYNAPS – A Journal of Professional Communication*, *24*, 41-50. http://hdl.handle.net/11250/2406059

Fjelland, R. (2020). Why general artificial intelligence will not be realized. *Humanities and Social Sciences Communications*, *7*(10), 1-9. https://doi.org/10.1057/s41599-020-0494-4

Fjelland, R. (2022). Teaching Philosophy of Science to Science Students: An Alternative

Approach. *Studies in Philosophy and Education*, *41*(2), 243-258. https://doi.org/10.1007/s11217-021-09802-8

Fraser, H., Parker, T., Nakagawa, S., Barnett, A., & Fidler, F. (2018). Questionable research practices in ecology and evolution. *PloS one*, *13*(7), e0200303. https://doi.org/10.1371/journal.pone.0200303

Galilei, G. (2016). The assayer. In *The Controversy on the Comets of 1618* (pp. 151-336). University of Pennsylvania Press.

Gandevia, S. (2018). Publication pressure and scientific misconduct: why we need more open governance. *Spinal Cord*, *56*(9), 821-822. https://doi.org/10.1038/s41393-018-0193-9

Gelman, A., & Loken, E. (2013). The garden of forking paths: Why multiple comparisons can be a problem, even when there is no “fishing expedition” or “p-hacking” and the research hypothesis was posited ahead of time. *Department of Statistics, Columbia University*, *348*.

Gigerenzer, G. (1993). The superego, the ego, and the id in statistical reasoning. In G. Keren & C. Lewis (Eds.), *A handbook for data analysis in the behavioral sciences: Methodological issues* (pp. 311-339). Lawrence Erlbaum Associates, Inc.

Gigerenzer, G. (2018). Statistical rituals: The replication delusion and how we got there. *Advances in Methods and Practices in Psychological Science*, *1*(2), 198-218. https://doi.org/10.1177/2515245918771329

Godfrey-Smith, P. (2009). *Theory and Reality: An introduction to the philosophy of science*. University of Chicago Press.

Goodman, S. (2008). A Dirty Dozen: Twelve P-Value Misconceptions. *Seminars in Hematology*, *45*(3), 135-140. https://doi.org/https://doi.org/10.1053/j.seminhematol.2008.04.003

Greenland, S., Senn, S. J., Rothman, K. J., Carlin, J. B., Poole, C., Goodman, S. N., & Altman, D. G. (2016). Statistical tests, P values, confidence intervals, and power: a guide to

misinterpretations. *European journal of epidemiology*, *31*(4), 337-350. https://doi.org/10.1007/s10654-016-0149-3

Hanfstingl, B. (2019). Should we say goodbye to latent constructs to overcome replication crisis or should we take into account epistemological considerations? *Frontiers in psychology*, *10*, 1949. https://doi.org/10.3389/fpsyg.2019.01949

Heidegger, M. (1977). *Basic Writings* (D. F. Krell, Ed.). Harper Perennial Modern Thought.

Heidegger, M. (2010). *Being and Time*. State University of New York Press.

Husserl, E. (1970). *The Crisis of European Sciences and Transcendental Phenomenology: An introduction to phenomenological philosophy*. Northwestern University Press.

Kalinowski, P., Fidler, F., & Cumming, G. (2008). Overcoming the inverse probability fallacy: A comparison of two teaching interventions. *Methodology*, *4*(4), 152-158. https://doi.org/10.1027/1614-2241.4.4.152

Kant, I. (2000). *Critique of the Power of Judgment*. Cambridge University Press.

Koyré, A. (1943). Galileo and Plato. *Journal of the History of Ideas*, *4*(4), 400-428. https://doi.org/10.2307/2707166

Krueger, J. I., & Heck, P. R. (2017). The heuristic value of p in inductive statistical inference. *Frontiers in psychology*, *8*, 908. https://doi.org/0.3389/fpsyg.2017.00908

Lyotard, J.-F. (1984). *The Postmodern Condition: A report on knowledge*. Manchester University Press.

Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research: How often do they really occur? *Perspectives on Psychological Science*, *7*(6), 537-542. https://doi.org/10.1177/1745691612460688

Meehl, P. E. (1997). The problem is epistemology, not statistics: Replace significance tests

by confidence intervals and quantify accuracy of risky numerical predictions. In L. L. Harlow, S. A. Mulaik, & J. H. Steiger (Eds.), *What if there were no significance tests* (Vol. 1, pp. 393–425.). Erlbaum.

Moran, D. (2012). *Husserl's Crisis of the European Sciences and Transcendental Phenomenology: An introduction*. Cambridge University Press.

Morawski, J. (2019). The replication crisis: How might philosophy and theory of psychology be of use? *Journal of Theoretical and Philosophical Psychology*, *39*(4), 218-238. https://doi.org/10.1037/teo0000129

Morawski, J. (2021). How to True Psychology's Objects. *Review of General Psychology*. https://doi.org/10.1177/10892680211046518

Muijs, D. (2004). Introduction to quantitative research. In *Doing quantitative research in education with SPSS* (pp. 1-12). Sage Publication Ltd.

Nuijten, M. B., Hartgerink, C. H., Van Assen, M. A., Epskamp, S., & Wicherts, J. M. (2016). The prevalence of statistical reporting errors in psychology (1985–2013). *Behavior research methods*, *48*(4), 1205-1226. https://doi.org/10.3758/s13428-015-0664-2

Perezgonzalez, J. D. (2017). Commentary: the need for Bayesian hypothesis testing in psychological science. *Frontiers in psychology*, *8*, 1434. https://doi.org/10.3389/fpsyg.2017.01434

Porter, T. M. (2020). *Trust in Numbers*. Princeton University Press. https://doi.org/10.1515/9780691210544

Proulx, T., & Morey, R. D. (2021). Beyond statistical ritual: theory in psychological science. *Perspectives on Psychological Science*, *16*(4), 671-681. https://doi.org/10.1177/17456916211017098

Reiss, J., & Sprenger, J. (2020). Scientific objectivity. In E. N. Zalta (Ed.), *Stanford Encyclopedia of Philosophy* (Winter ed.). Stanford.

Romeijn, J.-W. (2017). Philosophy of statistics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy* (Spring 2017 ed.).

Shrout, P. E., & Rodgers, J. L. (2018). Psychology, science, and knowledge construction: Broadening perspectives from the replication crisis. *Annual review of psychology*, *69*, 487-510. https://doi.org/10.1146/annurev- psych- 122216- 011845

Simmons, J. P., Nelson, L. D., & Simonsohn, U. (2011). False-Positive Psychology:Undisclosed Flexibility in Data Collection and Analysis Allows Presenting Anything as Significant. *Psychological Science*, *22*(11), 1359-1366. https://doi.org/10.1177/0956797611417632

Sober, E. (2002). Intelligent design and probability reasoning. *International Journal for Philosophy of Religion*, *52*(2), 65-80. https://doi.org/10.1023/A:1019579220694

Sober, E. (2008). *Evidence and Evolution: The logic behind the science*. Cambridge University Press.

Szucs, D., & Ioannidis, J. (2017). When null hypothesis significance testing is unsuitable for research: a reassessment. *Frontiers in human neuroscience*, *11*, 390. https://doi.org/10.3389/fnhum.2017.00390

Wittgenstein, L. (1969). *On Certainty.* (G. E. M. Anscombe & G. H. von Wright, Eds. 41 ed.). HarperCollins Publishers Inc.

Ziliak, S. T., & McCloskey, D. N. (2009, August 3rd). *The Cult of Statistical Significance,* the Joint Statistical Meetings, Washington, DC.