Theoretical and methodological foundations of experimental psychological research

Kumulative Inaugural-Dissertation zur Erlangung der Doktorwürde

der Fakultät für Humanwissenschaften

der Universität Regensburg

vorgelegt von

Roland Mayrhofer

aus Regensburg

Die Arbeit entstand in Betreuung durch

Prof. Dr. Christof Kuhbandner

(Lehrstuhl für Psychologie VI, Pädagogische Psychologie)

Regensburg 2024

Erstgutachter: Prof. Dr. Christof Kuhbandner

Zweitgutachter: Prof. Dr. Peter Fischer

Obergutachter: Prof. Dr. Hans Gruber

Contents

Preface

In this thesis, three studies are presented that explore how the general structure of experimental psychology affects specific research and experiments and, consequently, the nature of the knowledge produced. All three studies were published in peer-reviewed journals. They are reproduced in the accepted version. The contributions of the co-authors to the three studies are given below on page 5. All references were combined into one reference list at the end of this thesis, the numbering of the original chapter headings was adjusted or removed, and the in-text references were harmonized according to the APA 7 recommendations. Other than that, the manuscripts of the three studies remain unchanged.

Contributions

Study 1 – The Principle of Inversion: Why the Quantitative-Empirical Paradigm Cannot Serve as a Unifying Basis for Psychology as an Academic Discipline

Idea: Roland Mayrhofer, Fabian Hutmacher

Writing: Roland Mayrhofer

Revision: Roland Mayrhofer, Fabian Hutmacher

Study 2 – The Practice of Experimental Psychology. An Inevitably Postmodern Endeavor

Idea: Roland Mayrhofer, Christof Kuhbandner, Corinna Lindner

Writing: Roland Mayrhofer

Revision: Roland Mayrhofer, Christof Kuhbandner, Corinna Lindner

Study 3 – Re-examining the testing effect as a learning strategy: the advantage of retrieval practice over concept mapping as a methodological artifact

Idea: Roland Mayrhofer, Christof Kuhbandner

Design: Roland Mayrhofer, Christof Kuhbandner

Data Collection: Roland Mayrhofer, Katja Frischholz

Analysis: Roland Mayrhofer, Christof Kuhbandner, Katja Frischholz

Writing: Roland Mayrhofer

Revision: Roland Mayrhofer, Christof Kuhbandner, Katja Frischholz

List of Figures

1. Introduction

1.1 Background and current state of research: psychology as an academic discipline

Science is the attempt to generate reliable knowledge about the world in a systematic form by describing, classifying, and explaining phenomena, by discovering universal laws, that allow phenomena to be predicted, manipulated, and thereby controlled (e.g., Bonnay, 2018). Although this statement may seem easy to understand, none of these points are trivial, as the extremely complex – and also contradictory – discussions and results of the relevant philosophical subdisciplines show. All these levels are brought together in philosophy of science with the aim of making this complexity both navigable and usable for the needs and purposes of science as a human activity – and thus also for the scientists who actually carry out science (e.g., Machamer & Silberstein, 2002; Psillos, 2007; Staley, 2014). There are also more specific theories for individual areas of science, such as the natural sciences (e.g., Godfrey-Smith, 2021; Zinman, 2000), social sciences (e.g., Cartwright & Montuschi, 2014; Hollis, 1994; Kincaid, 2012) and humanities (e.g., Leezenberg & De Vries, 2018), but also for many individual scientific disciplines.

So how does this huge amount of knowledge affect psychology as an academic discipline and its endeavor to generate reliable knowledge about mind and behavior? There is a substantial number of works that deal specifically with the philosophy of psychology, but these are usually very broad questions that in some way affect psychology as a whole, such as general philosophical, usually ontic or epistemic, questions related to psychology – e.g., the nature of the mind and its relation to corporeality and biology – the scope of psychology, or ethical and social implications of its findings, especially in applied subdisciplines such as clinical or educational psychology (e.g., Bermúdez, 2006; Botterill & Carruthers, 1999; Bunge & Ardila, 1987; Robins et al., 2020; Thagard, 2007; Walsh et al., 2014; Weiskopf & Adams, 2015). Considering that a large part of philosophy of science revolves precisely around how reliable knowledge can be generated, it is surprising for two reasons that this question is hardly addressed: First, the way in which knowledge is generated is obviously of crucial importance for any scientific discipline, precisely because it is the goal of why the discipline is practiced in the first place. Second, especially in a discipline such as psychology, where a great deal of emphasis is placed on methodology, one would expect the relationship between the methods used, the object of research, and the results generated to be the subject of intense discussion.

Although individual substantive topics are discussed intensively – and controversially – including questions of methodology, there are hardly any works that deal with the role of methodology on a fundamental, abstract level in the sense of the philosophy of science. There are two possible reasons for this: On the one hand, it is possible that this issue has already been fully and satisfactorily resolved, which is why there is simply no need to question the existing system or seek changes. On the other hand – and this is the idea that is examined more closely in this study – it is also possible that there is still little awareness in psychology of the importance and also the explosive power of these connections, with enormous consequences for how psychology functions as a scientific discipline and what results, i.e., what knowledge it produces.

As already mentioned, there is a plethora of works that deal with methodology in psychology, be it detailed chapters in introductions (e.g., Gerrig, 2012; Lilienfeld et al., 2015; Myers & Dewall, 2015) or textbooks dedicated solely to this topic (e.g., Coolican, 2019; Howell, 2010; Shaughnessy et al., 2015). However, and this is the important point here, they deal primarily with the mechanics and functioning of a particular type of methodology, namely quantitative and experimental methods and techniques. This is certainly not surprising as these are the methods that are dominant in scientific psychology, especially in research, and shape its character to a large extent. The reasons that led to the dominance of the quantitative-experimental paradigm have so far been largely unexplored by the history of science. Although the answer to this question lies outside the scope of this study, the reasons would be of interest not only from a historical perspective but also from a philosophy of science perspective. For if this dominance of the quantitative-experimental paradigm could be fully explained by philosophy of science – i.e., by great successes in generating reliable knowledge and discovering universal laws with corresponding theoretical justifications – this would indicate that precisely this approach is fruitful and promising. From this extensive and complex topic, this study focuses on one aspect, namely the experiment as a method for generating knowledge in psychology, and examines whether and to what extent this is actually the case or whether there is also the danger that the use of experiments can, on the contrary, also generate unreliable or even false “knowledge”.

1.2 Objectives: theoretical and empirical analyses of the experiment in psychology

There is an extensive body of literature on the experiment as a method, especially in physics, from a philosophy of science perspective (e.g., Franklin, 1986 & 1990; Franklin & Perovic, 2023; Galison, 1987; Van Fraassen, 1980). However, there is a lack of comparable analyses for psychology in which the quantitative-experimental paradigm, and especially the experiment, as a method is theoretically illuminated and concrete experiments from psychology are analyzed from such a philosophy of science perspective. Focusing on the experiment as a research method in psychology, this study aims to close this gap by presenting and discussing two main arguments: First, it is shown how certain basic methodological and conceptual assumptions and presuppositions – i.e., the quantitative-experimental paradigm – which seem so self-evident in research practice, manifest themselves as a virtually fixed frame of reference within which psychological research takes place. Empirical examples are then used to show how these assumptions affect the research, the operationalization, the data collection, the results and ultimately their interpretation. Second, on this basis, the argument is put forward that both a more comprehensive and in-depth knowledge of abstract philosophy of science backgrounds is an indispensable component for the concrete implementation of experimental psychological research. This argument complements and extends the work of Haig (2014, 2018a & 2018b), who has already pointed out that in psychology – generally speaking – philosophy of science plays only a very minor role, which in turn can lead to various problems in research and the knowledge generated. Haig's main focus lies on epistemological aspects, how theories are generated, with particular emphasis on the hypothetico-deductive method, the inductive method, and reasoning, as well as the role of models in explaining phenomena. These fundamental aspects are complemented in the present study not only by specifically examining the experiment as an important research method in psychology from a philosophy of science perspective, but also by verifying the conclusions thus drawn on the basis of concrete experiments by means of an empirical investigation.

These two main objectives of the present analysis are developed on the basis of three studies: Study 1 begins by looking at psychology as an academic discipline as a whole. Jüttemann's (1983) observation that psychology, in contrast to most other scientific disciplines, is characterized less by its object of research rather than by its method, namely a quantitative-experimental approach that tends to ossify into a rather rigid paradigm, is analyzed in detail. The result that this quantitative-experimental paradigm actually shapes psychology – i.e., how research is conducted and the knowledge thus generated is influenced accordingly – forms the basis for Study 2.

There, in Study 2, the focus lies specifically on the experiment as a research method, which is analyzed against the backdrop of postmodern approaches from a philosophy of science perspective. This study demonstrates first at the theoretical level that the experiment does not – as is often assumed – simply faithfully reproduce the phenomenon of interest, but rather *re*-constructs it according to the assumptions that the researcher has implicitly or explicitly made. In combination with the result from Study 1, according to which the quantitative-experimental paradigm often determines the framework within which one thinks, and which thus preconfigures the approach to a phenomenon, it thus becomes clear that this reconstruction of a phenomenon therefore also takes place along such lines. This reasoning is then illustrated by two empirical findings. On the one hand, there is a very narrowly defined topic from basic research in cognitive psychology, namely the capacity of the visual working memory. On the other hand, a very broad topic was chosen that deals with a frequently and controversially discussed fundamental question of philosophy, namely the problem of free will. Both topics have in common that they deal with questions that at first glance appear to be clearly answerable – by a concrete number or a yes or no – and that were examined from a psychological perspective by means of a quantitative-experimental approach. They are therefore suitable for examining whether and to what extent certain more or less fixed presuppositions and thinking along certain patterns, as manifested by the outstanding importance of the experiment as a method of psychology, affect the type of research, the results, and their interpretation, as well as how this is communicated to the public.

In Study 3, the influence of the quantitative-experimental paradigm, within which psychological research often takes place, is empirically investigated. A prominent experimental study by Karpicke and Blunt (2011), which found that, aiming at meaningful learning, retrieval practice produces more learning than elaborative studying with concept mapping, is subjected to a methodological critique. This criticism is based on the assumption that Karpicke and Blunt's (2011) study is already so preconfigured due to certain assumptions inherent in the quantitative-experimental paradigm that the approach to the phenomenon of interest, namely to compare the efficiency of different learning strategies with regard to meaningful learning – i.e., an extensive and complex theoretical construct – is thereby affected. In other words, the decision to examine such a broad and complex topic as meaningful learning (e.g., Ausubel, 1963 & 1968; Agra et al., 2019; Cavallo, 1996; Jonassen & Strobel, 2006; Mayer, 2002; Shuell, 1990; Vallori, 2014) using simple quantitative parameters in an experimental setting and to draw broad conclusions from it reflects exactly the kind of thinking described in Study 1 and Study 2. Study 3 therefore pursues two objectives. First, on the level of the topic “meaningful learning”, it is investigated whether the observed advantage of retrieval practice over concept mapping reflects a biased experimental setting or is actually due to the particular learning strategy. Second, on a meta-experimental level, so to speak, it is also possible to examine whether the characteristics of the experiment as a research method, as identified in Study 1 and Study 2, are valid and whether modifications to the setting can lead to fundamentally different results.

1.3 Methodological considerations: applying philosophy of science to psychology

Before presenting Studies 1-3, some terms and methodological aspects are first clarified: Although non-quantitative experiments are of course also possible, experiments in psychology are generally quantitative. This is why this study uses the term "quantitative-experimental paradigm", and, unless otherwise specified, the term "experiment" refers to quantitative experiments. The questions of whether non-quantitative experiments have different characteristics than quantitative ones, or whether non-experimental quantitative studies have the same or comparable characteristics as quantitative experiments, would be of interest from a philosophy of science perspective but lie outside the scope of this study. The experiment as a research method in psychology is the central focus of this study and the term "experiment" is therefore analyzed in detail in Studies 1 and 2. In addition, statements are made here about psychology as a scientific discipline, but it is not implied that this applies to the whole of academic psychology or that psychology is a monolithic block in which there are no differences. Instead, the statements made here are intended as tendencies that do, however, affect larger parts of psychology. This aspect is examined in more detail in Study 1.

A third key concept in this study is "phenomenon", which, borrowing from Craver and Tabery (2023), Glennan (1996 & 2002), and Kauffman (1970), is understood here as something that can be experienced – either objectively perceived or subjectively experienced – and is perceived as a coherent whole. This intentionally extremely broad definition thus encompasses – in addition to material occurrences – all occurrences that psychology deals with, which already hints at another basic theme of Study 1, namely the object of investigation of psychology.

Furthermore, two fundamental caveats need to be clarified, which are not the intended aims of this study. The first caveat is that this study is not – as is often the case in philosophy of science – an investigation into various general topics related to science *per se*, such as causality, theorizing, or the use of models. Instead, the focus here is on how causality is considered in experimental psychological research, how theories and models are used and the like, i.e., how said aspects are implemented in actual research practice.

The second caveat is that the present study is not prescriptive, i.e., in the sense that it lays down rules or makes recommendations as to how psychological research should (not) be conducted, as was done for science in general in logical positivism or empiricism (e.g., Achinstein & Barker, 1969; Richardson & Uebel, 2007; Uebel & Limbeck-Lilienau, 2022) and in critical rationalism (e.g., Keuth, 2005; Rowbottom, 2011). There are two reasons for this:

First, universal prescriptive systems are, as the examples of logical empiricism or critical rationalism just mentioned show, generally too abstract and too rigid to adequately describe an extensive and complex undertaking such as science (probably most pointedly argued by Feyerabend, 1975; see also Kantorovich, 1988; Niaz, 2020). The ideas of philosophy of science schools of thought, which prescribed very rigorous rules for scientific research practice, above all logical empiricism and critical rationalism, are now generally regarded as inadequate and obsolete (Bird, 2013). Second, and this reason is crucial here, such universal prescriptive systems presuppose certain axioms which, if followed, shape science and the approach to the investigated phenomena in a certain way that reflects precisely these a priori assumed axioms in a circular, self-confirming and self-validating way, so to speak, by excluding other approaches – which could lead to other results – from the very outset. Since the present study examines this very mechanism, it is therefore necessary that this potentially problematic procedure – in reverse, as it were – is not repeated here.

Nevertheless, the present study is also necessarily based on certain presuppositions, which are set out below. First, it is assumed here that science is not an isolated endeavor that stands alone, obeying only its own rules that enable the optimal generation of knowledge (e.g., with different focuses and emphases, Agassi, 1981; Bourdieu, 1984; Bucchi, 2002; Feyerabend, 1975; Fleck, 1935/1980; Kuhn, 1962; Latour, 1999; Matthew, 2005). In this study, it is less important how exactly science functions as a social system – in sociology and philosophy of science, different ways of functioning and mechanisms are postulated and discussed – but rather that science is also a social system whose characteristics shape how the object of research is viewed and approached, the research methods, and thus also the results and their interpretation. Interestingly, in this context Fleck (1935/1980) has already shown for bacteriology and Feyerabend (1975) for astronomy how much the perception as well as the investigation and interpretation of phenomena – even and especially of seemingly "objective" and "naturally" occurring phenomena – depends on various different factors. These include fundamental, general ideas about the nature of the world, the way in which science is understood, concrete existing theories, both about the phenomenon of interest itself and theories about other phenomena that may be related to it, the instruments used, the aims of the investigation as well as the narrower and wider social context of the researchers.

Second, and closely related to the first presupposition, another fundamental dimension, explored in particular by Feyerabend (1975), is the use of language – in the broadest sense – which additionally permeates all these aspects in two ways. The relationship between language, thought, cognitive possibilities, and the nature of the world is extremely complex and has therefore long been the subject of extensive and controversial debate in various disciplines (see, e.g., Devitt & Hanley, 2006; Lycan, 2019; Platts, 1997). Although this topic naturally goes far beyond the scope of this study, some aspects are nevertheless of fundamental importance here. The first aspect may seem trivial but nevertheless forms the basis for any kind of science – including the present study itself – because without some form of language no communication and thus obviously no science would be possible. However, this connective element also harbors the risk of misunderstandings and miscommunication. On the one hand, it is possible that different terms are used for the same concept, so that problems may arise at the level of language that do not exist at the factual level. On the other hand, there is the possibility that the same term is understood to mean different concepts, which is why the apparent agreement on the linguistic level may conceal differences on the factual level – even without those involved being aware of this.

Furthermore, a third possible problem arises here, namely the connection between an observable phenomenon and cognitive mechanisms that are usually not observable in psychology, which must be translated into concepts. This is not only necessary in order to be able to communicate about it, but especially in order to be able to form a concept of the corresponding phenomenon or concept in the first place, because things for which humans have no concept are generally also not perceived. However, when phenomena and concepts are captured in a linguistic term, this also means that they are already linguistically and mentally pre-structured and possibly loaded with content that may not actually exist or may exist in a different form. Without elaborating on this enormously complex philosophical problem of language here, it should be noted for the purposes of the present study that science – especially when it works with immaterial constructs, as is very often the case in psychology – is highly dependent on the language used to adequately describe the respective concepts and phenomena. Against this background, there is a risk that the terms used – and consequently also the operationalization of the construct of interest in the experiment, the results and their interpretation – become blurred and are used merely as empty phrases which, although they suggest a precise content, are in reality unclear or ambiguous.

Finally, this leads to a last point that is relevant to the present study, namely the way in which findings generated by experimental psychological research are justified as knowledge. As Feyerabend (1975) in particular has shown with examples from astronomy and is also implied by the fact that science is embedded in a social context, the persuasive power of scientific research does not only consist in its results meeting certain factual and objective criteria and being convincing, for example, through agreement with existing knowledge, internal consistency, or predictive power. Instead, a variety of other “non-scientific”, “subjective” factors such as personal dispositions, cognitive, emotional, motivational, and social factors, or intuition play a role in the assessment of the results of scientific research – just as they do in non-scientific questions. At this point, the role of language becomes relevant again, as these points are conveyed through language and thus the problems mentioned above can also occur, so that the investigation of a phenomenon in the experiment as well as results and their interpretation are subject to a variety of factors that can affect and even compromise the actual intention – namely the generation of reliable knowledge.

The aspects set out in this brief outline are examined in more detail below on the basis of three studies, before their results are then brought together in a synthesis along the lines described here.

2. Peer-Reviewed Studies

2.1 Study 1 – The Principle of Inversion: Why the Quantitative-Empirical Paradigm Cannot Serve as a Unifying Basis for Psychology as an Academic Discipline

This is a pre-copy-editing, author-produced version of an article published 2020 in Frontiers in Psychology (Cognitive Science) following peer review. The official citation that should be used in referencing this material is: Mayrhofer, R., & Hutmacher, F. (2020). The Principle of Inversion: Why the Quantitative-Empirical Paradigm Cannot Serve as a Unifying Basis for Psychology as an Academic Discipline. *Frontiers in Psychology, 11*, 596425. https://doi.org/10.3389/fpsyg.2020.596425.

Keywords: self-perception of psychology, principle of inversion, methods in psychology, operationalism, definitions of psychology

Introduction

In the English-speaking world, as well as in the international academic discourse and many other languages, the term “science” or appropriate translations refer only to a certain area of knowledge, namely the natural and the social sciences, thus excluding what is usually referred to as humanities or *Geisteswissenschaften* (Szostak, 2004).

The history and philosophy of science shows that the sciences rely heavily on the so-called Scientific Method, a set of theoretical and methodological principles which consists, in essence, of observing, formulating hypotheses and testing these hypotheses in experiments, in order to discover general laws. In doing so, knowledge is generated by relying on empirical evidence, which in turn expresses directly observable phenomena in terms of quantitative data. Excluding the extremely complex discussions about advantages, drawbacks, and alternatives to the Scientific Method (e.g., Gower, 1997; Nola & Sankey, 2014; Andersen & Hepburn, 2015), it is important to note that quantitative-empirical methods and thinking according to the Scientific Method dominate the sciences (Haig, 2014; Nola & Sankey, 2014; Sankey, 2014) and also psychology (Garber, 2019; Haig, 2019; Toomela, 2020).

By contrast, the humanities rely much less on quantitative – let alone experimental – methods, although these are employed nonetheless when appropriate. Instead of producing and gathering empirical data, the humanities characteristically approach their subject matter from a descriptive, interpretive, and hermeneutical understanding whose historical and comparative angles cannot be conveniently summarized by a single term (Watanabe, 2010; Bem & De Jong, 2013; Leezenberg & De Vries, 2018).

Psychology as an Academic Discipline and the Principle of Inversion

Psychology is usually portrayed as the study of the human mind and behavior, although this nomenclature does not make it entirely clear what it actually designates.

In ancient Greek, "*psyche*" ("ψυχή") encompasses a variety of meanings (Claus, 1983) such as stream of air, breath of life, substance of life (in an ontological sense), spirit, mind, soul, personality, consciousness, self, or even ghost (of the dead). Although these terms describe a semantic field with the underlying term "life," this does not tell us what psychologists actually investigate. The term "psychology," in the sense of "study of the soul," originated in the early modern era and was employed mostly for topics which would be categorized to be part of philosophy today. Nevertheless, the founders of modern academic psychology with its predominantly experimental and empirical orientation in the nineteenth century – most notably Wilhelm Wundt – continued to use this term. Further historical research (such as Russo Krauss, 2019) might explain why "*psyche*" was retained as term for their subject matter, thus clarifying the conceptual ideas behind actual research.

This heterogeneity of "psyche" is mirrored in the different psychological subdisciplines, such as cognitive, social, or biological psychology and the many branches of applied psychology, such as educational, organizational, or clinical psychology. And while mind, personality, consciousness, and self are familiar terms in psychology, it is clear that other aspects of the Greek "*psyche*" such as the physical properties of breath are not part of the discipline.

In short, psychology investigates many aspects of human existence – but then how does psychology differ from, say, anthropology, sociology, or history? Wherein lies the unity of psychology as academic discipline?

A widely-used textbook (Gerrig, 2012, p. 2) gives the following answer: Many psychologists seek answers to the fundamental question: “What is human nature?” This question is pursued by looking at processes that occur within individuals, and thus psychology is defined as the scientific study of the behavior of individuals and their mental processes (see also, e.g., Lilienfeld et al., 2015; Myers & Dewall, 2015 for similar conceptualizations). However, the main elements of this train of thought – human nature, the individual in its entirety, and a scientific approach – reveal that it is not a trivial matter to state precisely what psychology is actually about.

First, it seems debatable that Gerrig’s attempt to subsume the subject matter of psychology under the umbrella term “human nature” is really more precise than hazy concepts such as “*psyche*” or “soul” which contemporary psychology has dismissed as too vague (Haaga, 2004; Henriques, 2004; Lilienfeld, 2004). Whereas academic psychologists might argue that they are not interested in such a hazy concept but rather in specific topics such as emotions, neurobiology, or education, all these concepts revolve around the human mind and behavior. Therefore, psychology does indeed have some kind of common theme or center – but this center is so vague that it cannot act as focal point or provide the same clear framework as the subject matters of other disciplines. By contrast, physics is also very diverse, possibly even more so than psychology, but its subject matter is clearly defined as matter and the related phenomena of energy, space and time.

Second, Gerrig’s assumption that the answer to the question about “human nature” can be found “by looking at processes that occur within individuals” is not self-evident. Simply put, focusing on the individual is problematic because many—if not all—of the individuals’ intrinsic processes are inextricably intertwined with larger social, societal, or historical contexts (e.g., Agassi, 1977; Margolis, 1995 & 2008). In other words, the behavior of individuals and their mental processes are shaped by outside contextual and societal factors, which may vary over time. The failure to take variability into consideration might underestimate the complexity of mental processes and give the impression that “human nature” is more hard-wired and less context-dependent than it actually is.

All of this shows that the exact subject matter of psychology is hard to pinpoint or to distinguish from other disciplines which also deal with behavior and mental processes, such as anthropology, history, cultural and literary studies, or philosophy. Nevertheless, Gerrig’s definition contains an element which is crucial for the self-conception of psychology as discipline, namely the emphasis on “scientific study.” Similarly, the APA Dictionary of

Psychology explicitly emphasizes "observation, experimentation, testing, and analysis" (VandenBos, 2015, p. 860) as characteristic research methods, echoing the Scientific Method.

As early as 1983, Jüttemann pointed out that the common factor underlying the various branches and areas of research in psychology is not characterized thematically but rather by a common method and methodology, namely the rather strict – and sometimes dogmatic – adherence to the Scientific Method. He termed this "the principle of inversion": in other disciplines method and methodology are aligned with the respective subject matter, but in psychology this principle is inverted (see also Royce, 1961; Michell, 1997; Summers, 2012).

Jüttemann's astute observation has two interesting consequences: First, there is a stark contrast between the rather strict methodological requirements and the very broad and often hazy thematic content of psychological research. Therefore, Jüttemann concludes ironically, "everything done by psychologists who employ the nomological methodology in their research must count as academic psychology" (Jüttemann, 1983, p. 34, translated). Second, this modus operandi differs from other academic fields which either have clearly defined subject matters or employ much less rigid methodologies.

The Scientific Method originated long before the institutionalization of psychology as an academic discipline in the late nineteenth century. Moreover, it is closely associated with the natural sciences, especially physics. Thus, psychology is dominated by a method which is neither unique to nor was developed within the own academic field. By contrast, methods which were developed within and specifically for the framework of psychology such as psychoanalysis or introspection are relegated to the fringes of academic psychology.

Jüttemann argues that the rigid methodology has far-reaching consequences for the very nature of psychology and criticizes the resulting "research operationalism," i.e., the fact that the subject matters of certain areas – such as stress – are only represented by phenomena and procedures which conform to operationalizations according to the Scientific Method. But by reducing complex phenomena to easily quantifiable laboratory procedures the concepts in question – e.g., stress – lose their original meaning. In essence, this means that by solely using the Scientific Method to investigate psychological phenomena (such as stress) we do not learn much about these phenomena as such. Rather, we transform them into something else which can be quantified and measured, meaning that highly complex phenomena are simplified in order to make them quantifiable (see also Hibberd, 2019; Mayrhofer et al., 2021). "Stress" is a concrete example because the concept of stress "is deeply intertwined with the constituents of modern identity" (Hutmacher, 2019a, p. 181) and therefore extremely complex, although research on stress highly relies on quantifiable parameters.

Whereas this operationalizing of psychological concepts might be appreciated as a more precise specification, it also goes hand in hand with a narrowing of real-life phenomena. It is possible that stripping complex phenomena down to their – supposedly! – bare bones reveals their core mechanisms. But more often we lose important aspects during this process, thereby missing the opportunity to understand something in its entirety. In other words, the professed exactitude and the desire to uncover the fundamental mechanisms of mental and behavioral phenomena by employing a quantitative-empirical paradigm within the very framework of the Scientific Method inadvertently misses important aspects.

To put it differently, it is by no means self-evident that studying internal – mental or behavioral – processes according to a certain predefined method will yield the desired results or that other methods might not provide more or a different kind of insight. Furthermore, even if we gain some knowledge by applying the Scientific Method, it is not an evident conclusion that this will also tell us something about human nature.

Discussion

Where does this leave academic psychology? Thematically, psychology is a very colorful picture of different subject matters, whose interconnectedness is often rather tenuous and does not display a strong cohesiveness while circling around "human nature" as a hazy center of gravity. However, this vibrant mixture is hidden behind a veneer of uniformity, which manifests itself in the strict adherence to the quantitative-empirical method. This uniformity certainly conveys strength because of its methodological rigor and scientific respectability. However, simultaneously it hampers psychology by preventing it from exploring other avenues which might yield additional insight into mental and behavioral processes or even human nature.

The idea of a one-stop method is problematic for two reasons: First, psychology as a field is wide and diverse. Second, the specific mental and behavioral phenomena – such as stress – are hard to define precisely (Zagaria et al., 2020). Therefore, applying something seemingly precise such as the Scientific Method is inherently at odds with trying to understand such hard-to-grasp, complex phenomena. In short, the quantitative-empirical method cannot serve as a unifying basis for psychology as an academic discipline because it misses important dimensions of "human nature."

We believe that postmodern approaches, which were specifically developed to describe the complexities and ambiguities of modern societies, may offer a way out of this dilemma, although here we can only give a brief sketch: Postmodern approaches recognize and emphasize that a certain phenomenon may be understood by using different methods. Seemingly different phenomena and/or approaches often point into the same direction, although from different perspectives (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015). This does not mean that there is no "truth" in psychology or that we cannot approach this truth (Holtz, 2020). Rather, the strength of a postmodern mindset lies in the ability to describe and to comprehend very complex phenomena without watering them down.

Therefore, such approaches will probably expand both the range and the explanatory power of psychology. As mental and behavioral processes tend to be innately fuzzy, any investigation of these phenomena must take this fuzzy nature into account. This is of course no plea to abandon the quantitative-empirical methods as they have revealed many interesting aspects of the psyche. The human mind and behavior are diverse – so why should our methods for investigating them not be equally diverse? Although largely outside the "scientific mainstream," there are other schools of thought in psychology which operate on the basis of different concepts of science, such as psychoanalysis (Bazan, 2018), humanistic (Warmoth, 1998), constructivist (Lincoln & Hoffman, 2019), or phenomenological (Langdridge, 2007) psychology. Taking their approaches seriously is likely to turn academic psychology into a vibrant generator of relevant knowledge and to spark more light to the enigma we term "*psyche*".

2.2 Study 2 – The Practice of Experimental Psychology. An Inevitably Postmodern Endeavor

This is a pre-copy-editing, author-produced version of an article published 2021 in Frontiers in Psychology (Cognitive Science) following peer review. The official citation that should be used in referencing this material is: Mayrhofer, R., Kuhbandner, C., & Lindner, C. (2021). The Practice of Experimental Psychology. An Inevitably Postmodern Endeavor. *Frontiers in Psychology, 11*, 612805. https://doi.org/10.3389/fpsyg.2020.612805.

Abstract

The aim of psychology is to understand the human mind and behavior. In contemporary psychology, the method of choice to accomplish this incredibly complex endeavor is the experiment. This dominance has shaped the whole discipline from the self-concept as an empirical science and its very epistemological and theoretical foundations, via research practice and the scientific discourse to teaching. Experimental psychology is grounded in the scientific method and positivism, and these principles, which are characteristic for modern thinking, are still upheld. Despite this apparently stalwart adherence to modern principles, experimental psychology exhibits a number of aspects which can best be described as facets of postmodern thinking although they are hardly acknowledged as such. Many psychologists take pride in being "real natural scientists" because they conduct experiments, but it is particularly difficult for psychologists to evade certain elements of postmodern thinking in view of the specific nature of their subject matter. Postmodernism as a philosophy emerged in the 20th century as a response to the perceived inadequacy of the modern approach and as a means to understand the complexities, ambiguities, and contradictions of the times. Therefore, postmodernism offers both valuable insights into the very nature of experimental psychology and fruitful ideas on improving experimental practice to better reflect the complexities and ambiguities of human mind and behavior. Analyzing experimental psychology along postmodern lines begins by discussing the implications of transferring the scientific method from fields with rather narrowly defined phenomena – the natural sciences – to a much broader and more heterogeneous class of complex phenomena, namely the human mind and behavior. This ostensibly modern experimental approach is, however, per se riddled with postmodern elements: (re-)creating phenomena in an experimental setting, including the hermeneutic processes of generating hypotheses and interpreting results, is no carbon copy of “reality” but rather an active construction which reflects irrevocably the pre-existing ideas of the investigator. These aspects, analyzed by using postmodern concepts like hyperreality and simulacra, did not seep in gradually but have been present since the very inception of experimental psychology, and they are necessarily inherent in its philosophy of science. We illustrate this theoretical analysis with the help of two examples, namely experiments on free will and visual working memory. The postmodern perspective reveals some pitfalls in the practice of experimental psychology. Furthermore, we suggest that accepting the inherently fuzzy nature of theoretical constructs in psychology and thinking more along postmodern lines would actually clarify many theoretical problems in experimental psychology.

Keywords: postmodernism, experimental psychology, experiment, methodology, philosophy of science

Introduction

Postmodernism is, in essence, an attempt to achieve greater clarity in our perception, thinking, and behavior by scrutinizing their larger contexts and preconditions, based on the inextricably intertwined levels of both the individual and the society. Psychology also studies the human mind and behavior, which indicates that psychology should dovetail with postmodern approaches. In the 1990s and early 2000s, several attempts were made to introduce postmodern thought as potentially very fruitful ideas into general academic psychology (Jager, 1991; Kvale, 1992; Holzman & Morss, 2000; Holzman, 2006). However, overall they were met with little response.

Postmodern thoughts have been taken up by several fringe areas of academic psychology, e.g., psychoanalysis (Leffert, 2007; Jiménez, 2015; but see Holt, 2005), some forms of therapy and counseling (Ramey & Grubb, 2009; Hansen, 2015), humanistic (Krippner, 2001), feminist and gender (Hare-Mustin & Marecek, 1988; Sinacore & Enns, 2005), or cultural psychology (Gemignani & Peña, 2007).

In the following years, and continuing the so-called science wars of the 1990s (Segerstråle, 2000), several other attacks were launched against a perceived rise or even dominance of postmodern thought in psychology. Held (2007; see also the rebuttal by Martin & Sugarman, 2009) argued that anything postmodern would undermine rationality and destroy academic psychology. Similarly, postmodernism was identified – together with "radical environmentalism" and "pseudoscience" among other things – as a "key threat to scientific psychology" (Lilienfeld, 2010, p. 282), or as "inimical to progress in the psychology of science" (Capaldi and Proctor, 2013, 331). The following advice was given to psychologists: "We [psychologists] should also push back against the pernicious creep of these untested concepts into our field" (Tarescavage, 2020, p. 4). Furthermore, the term "postmodern" is even employed as an all-purpose invective in a popular scientific book by psychologist Steven Pinker (2018).

Therefore, it seems that science and experimental psychology on the one hand and postmodern thinking on the other are irreconcilable opposites. However, following Gergen (2001) and Holtz (2020), we argue that this dichotomy is only superficial because postmodernism is often misunderstood. A closer look reveals that experimental psychology contains many postmodern elements. Even more, there is reason to assume that a postmodern perspective may be beneficial for academic psychology: First, the practice of experimental psychology would be improved by integrating postmodern thinking because it reveals a side of the human psyche for which experimental psychology is mostly blind. Second, the postmodern perspective can tell us much about the epistemological and social background of experimental psychology and how this affects our understanding of the human psyche.

A Postmodern Perspective on Experimental Psychology

Experimental Psychology and the Modern Scientific Worldview

It lies within the nature of humans to try to find out more about themselves and their world, but the so-called Scientific Revolution of the early modern period marks the beginning of a new era in this search for knowledge. The Scientific Revolution, which has led to impressive achievements in the natural sciences and the explanation of the physical world (e.g., Olby et al., 1991; Henry, 1997; Cohen, 2015; Osterlind, 2019), is based on the following principle: to "measure what can be measured and make measurable what cannot be measured." This famous appeal – falsely attributed to Galileo Galilei but actually from the 19th century (Kleinert, 2009) – illustrates the two fundamental principles of modern science: First, the concept of "measurement" encompasses the idea that phenomena can be quantified, i.e., expressed numerically. Second, the concept of "causal connections" pertains to the idea that consistent, nonrandom relationships can be established between measurable phenomena. Quantification allows that relationships between phenomena can be expressed, calculated, and predicted in precise mathematical and numerical terms.

However, there are two important issues to be aware of. First, while it is not difficult to measure "evident" aspects, such as mass and distance, more complex phenomena cannot be measured easily. In such cases, it is therefore necessary to find ways of making these "elusive" phenomena measurable. This can often only be achieved by reducing complex phenomena to their simpler – and measurable! – elements. For instance, in order to measure memory ability precisely, possible effects of individual preexisting knowledge which introduce random variance and thus impreciseness have to be eliminated. Indeed, due to this reason, in many memory experiments, meaningless syllables are used as study material.

Second, it is not difficult to scientifically prove a causal relationship between a factor and an outcome if the relationship is simple, that is, if there is only one single factor directly influencing the outcome. In such a case, showing that a manipulation of the factor causes a change in the outcome is clear evidence for a causal relationship because there are no other factors which may influence the outcome as well. However, in situations where many factors influence an outcome in a complex, interactive way, proving a causal relationship is much more difficult. To prove the causal effect of one factor in such a situation the effects of all other factors – called confounding factors from the perspective of the factor of interest – have to be eliminated so that a change in the outcome can be truly attributed to a causal effect of the factor of interest. However, this has an important implication: The investigator has to divide the factors present in a given situation into interesting versus non-interesting factors with respect to the current context of the experiment. Consequently, while experiments reveal something about local causal relationships, they do not necessarily provide hints about the net effect of all causal factors present in the given situation.

The adoption of the principles of modern science has also changed psychology. Although the beginnings of psychology – as the study of the *psyche* – date back to antiquity, psychology as an academic discipline was established in the mid to late 19th century. This enterprise was also inspired by the success of the natural sciences, and psychology was explicitly modeled after this example by Wilhelm Wundt – the "father of experimental psychology" – although he emphasized the close ties to the humanities as well. The experiment quickly became the method of choice. There were other, more hermeneutic approaches during this formative phase of modern psychology, such as psychoanalysis or introspection according to the Würzburg School, but their impact on academic psychology was limited. Behaviorism emerged as a direct reaction against these perceived unscientific approaches, and its proponents emphasized the scientific character of their "new philosophy of psychology." It is crucial to note that in doing so they also emphasized the importance of the experiment and the necessity of quantifying directly observable behavior in psychological research. Behaviorism quickly became a very influential paradigm which shaped academic psychology. Gestalt psychologists, whose worldview is radically different from behaviorism, also relied on experiments in their research. Cognitive psychology, which followed, complemented, and partly superseded behaviorism, relies heavily on the experiment as a means to gain insight into mental processes, although other methods such as modeling are employed as well. Interestingly, there is a fundamental difference between psychoanalysis and humanistic psychology, which do not rely on the experiment, and the other abovementioned approaches as the former focus on the psychic functioning of individuals, whereas the latter focus more on global laws of psychic functioning across individuals. This is reflected in the fact that psychological laws in experimental psychology are established on the arithmetic means across examined participants – a difference we will elaborate on later in more detail. Today, psychology is the scientific – in the sense of empirical-quantitative – study of the human mind and behavior, and the experiment is often considered the gold standard in psychological research (e.g., Mandler, 2007; Goodwin, 2015; Leahey, 2017).

The experiment is closely associated with the so-called scientific method (Haig, 2014; Nola & Sankey, 2014) and the epistemological tenets philosophy of positivism – in the sense as Martin (2003); Michell (2003), and Teo (2018) explain – which sometimes exhibit characteristics of naïve empiricism. Roughly speaking, the former consists of observing, formulating hypotheses, and testing these hypotheses in experiments. The latter postulates that knowledge is based on sensory experience, that it is testable, independent of the investigator and therefore objective as it accurately depicts the world as it is. This means that in principle all of reality can not only be measured but eventually be entirely explained by science. This worldview is attacked by postmodern thinkers who contend that the world is far more complex and that the modern scientific approach cannot explain all of reality and its phenomena.

The Postmodern Worldview

Postmodern thinking (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015) has gained momentum since the 1980s, and although neither the term "postmodernism" nor associated approaches can be defined in a unanimous or precise way, they are characterized by several intertwined concepts, attitudes, and aims. The most basic trait is a general skepticism and the willingness to question literally everything from the ground up—even going so far as to question not only the foundation of any idea, but also the question itself. This includes the own context, the chosen premises, thinking, and the use of language. Postmodernism therefore has a lot in common with science's curiosity to understand the world: the skeptical attitude paired with the desire to discover how things really are.

Postmodern investigations often start by looking at the language and the broader context of certain phenomena due to the fact that language is the medium in which many of our mental activities – which subsequently influence our behavior – take place. Thus, the way we talk reveals something about how and why we think and act. Additionally, we communicate about phenomena using language, which in turn means that this discourse influences the way we think about or see those phenomena. Moreover, this discourse is embedded in a larger social and historical context, which also reflects back on the use of language and therefore on our perception and interpretation of certain phenomena.

Generally speaking, postmodern investigations aim at detecting and explaining how the individual is affected by societal influences and their underlying, often hidden ideas, structures, or mechanisms. As these influences are often fuzzy, contradictory, and dependent on their context, the individual is subject to a multitude of different causalities, and this already complex interplay is further complicated by the personal history, motivations, aims, or ways of thinking of the individual. Postmodernism attempts to understand all of this complexity as it is in its entirety.

The postmodern approaches have revealed three major general tendencies which characterize the contemporary world: First, societies and the human experience since the 20th century have displayed less coherence and conversely a greater diversity than the centuries before in virtually all areas, e.g., worldviews, modes of thinking, societal structures, or individual behavior. Second, this observation leads postmodern thinkers to the conclusion that the grand narratives which dominated the preceding centuries and shaped whole societies by providing frames of references have lost – at least partially – their supremacy and validity. Examples are religious dogmas, nationalism, industrialization, the notion of linear progress – and modern science because it works according to certain fundamental principles. Third, the fact that different but equally valid perspectives, especially on social phenomena or even whole worldviews, are possible and can coexist obviously affects the concepts of "truth," "reality," and "reason" in such a way that these concepts lose their immutable, absolute, and universal or global character, simply because they are expressions and reflections of a certain era, society, or worldview.

At this point, however, it is necessary to clarify a common misconception: Interpreting truth, reality, or reason as relative, subjective, and context-dependent – as opposed to absolute, objective, and context-independent – does naturally neither mean that anything can be arbitrarily labeled as true, real, or reasonable, nor, vice versa, that something cannot be true, real, or reasonable. For example, the often-quoted assumption that postmodernism apparently even denies the existence of gravity or its effects as everything can be interpreted arbitrarily or states that we cannot elucidate these phenomena with adequate accuracy because everything is open to any interpretation (Sokal, 1996), completely misses the point.

First, postmodernism is usually not concerned with the laws of physics and the inanimate world as such but rather focuses on the world of human experience. However, the phenomenon itself, e.g., gravity, is not the same as our scientific knowledge of phenomena – our chosen areas of research, methodological paradigms, data, theories, and explanations – or our perception of phenomena, which are both the results of human activities. Therefore, the social context influences our scientific knowledge, and in that sense scientific knowledge is a social construction (Hodge, 1999).

Second, phenomena from human experience, although probably more dependent on the social context than physical phenomena, cannot be interpreted arbitrarily either. The individual context – such as the personal history, motivations, aims, or worldviews – determines whether a certain behavior makes sense for a certain individual in a certain situation. As there are almost unlimited possible backgrounds, this might seem completely random or arbitrary from an overall perspective. But from the perspective of an individual the phenomenon in question may be explained entirely by a theory for a specific – and not universal – context.

As described above, the postmodern meta-perspective directly deals with human experience and is therefore especially relevant for psychology. Moreover, any discipline – including the knowledge it generates – will certainly benefit from understanding its own (social) mechanisms and implications. We will show below that postmodern thinking not only elucidates the broader context of psychology as an academic discipline but rather that experimental psychology exhibits a number of aspects which can best be described as facets of postmodern thinking although they are not acknowledged as such.

The Postmodern Context of Experimental Psychology

Paradoxically, postmodern elements have been present since the very beginning of experimental psychology although postmodernism gained momentum only decades later. One of the characteristics of postmodernism is the transplantation of certain elements from their original context to new contexts, e.g., the popularity of “Eastern” philosophies and practices in contemporary “Western” societies. These different elements are often juxtaposed and combined to create something new, e.g., new “westernized” forms of yoga (Shearer, 2020).

Similarly, the founders of modern academic psychology took up the scientific method, which was originally developed in the context of the natural sciences, and transplanted it to the study of the human psyche in the hope to repeat the success of the natural sciences. By contrast, methods developed specifically in the context of psychology such as psychoanalysis (Wax, 1995) or introspection according to the Würzburg School (Hackert & Weger, 2018) have gained much less ground in academic psychology. The way we understand both the psyche and psychology has been shaped to a great extent by the transfer of the principles of modern science, namely quantitative measurement and experimental methods, although it is not evident per se that this is the best approach to elucidate mental and behavioral phenomena. Applying the methods of the natural sciences to a new and different context, namely to phenomena pertaining to the human psyche, is a truly postmodern endeavor because it juxtaposes two quite distinct areas and merges them into something new – experimental psychology.

The postmodern character of experimental psychology becomes evident on two levels: First, the subject matter – the human psyche – exhibits a postmodern character since mental and behavioral phenomena are highly dependent on the idiosyncratic contexts of the involved individuals, which makes it impossible to establish unambiguous general laws to describe them. Second, experimental psychology itself displays substantial postmodern traits because both its method and the knowledge it produces – although seemingly objective and rooted in the modern scientific worldview – inevitably contain postmodern elements, as will be shown below.

The Experiment as Simulacrum

The term "simulacrum" basically means "copy," often in the sense of "inferior copy" or "phantasm/illusion." However, in postmodern usage "simulacrum" has acquired a more nuanced and concrete meaning. "Simulacrum" is a key term in the work of postmodern philosopher Jean Baudrillard, who arguably presented the most elaborate theory on simulacra (1981/1994). According to Baudrillard, a simulacrum "is the reflection of a profound ['real'] reality" (p. 16/6). Simulacra, however, are more than identical carbon copies because they gain a life of their own and become "real" in the sense of becoming an own entity. For example, the personality a pop star shows on stage is not "real" in the sense that it is their "normal," off-stage personality, but it is certainly "real" in the sense that it is perceived by the audience even if they are aware that it might be an "artificial" personality. Two identical cars can also be "different" for one might be used as a means of transportation while the other might be a status symbol. Even an honest video documentation of a certain event is not simply a copy of the events that took place because it lies within the medium video that only certain sections can be recorded from a certain perspective. Additionally, the playback happens in other contexts as the original event, which may also alter the perception of the viewer.

The post-structuralist – an approach closely associated with postmodernism – philosopher Roland Barthes pointed out another important aspect of simulacra. He contended that in order to understand something – an "object" in Barthes' terminology – we necessarily create simulacra because we "reconstruct [our italics] an 'object' in such a way as to manifest thereby the rules of functioning […] of this object" (Barthes, 1963/1972, p. 213/214). In other words, when we investigate an object – any phenomenon, either material, mental, or social – we have to perceive it first. This means that we must have some kind of mental representation of the phenomenon/object – and it is crucial to note that this representation is not the same thing as the "real" object itself. All our mental operations are therefore not performed on the "real" object but on mental representations of the object. We decompose a phenomenon in order to understand it, that is, we try to identify its components. In doing so, we effect a change in the object because our phenomenon is no longer the original phenomenon "as it is" for we are performing a mental operation on it, thereby transforming the original phenomenon. Identifying components may be simple, e.g., dividing a tree into roots, trunk, branches, and leaves may seem obvious or even "natural" but it is nevertheless us as investigators who create this structure – the tree itself is probably not aware of it. Now that we have established this structure, we are able to say that the tree consists of several components and name these components. Thus, we have introduced "new" elements into our understanding of the tree. This is the important point, even though the elements, i.e., the branches and leaves themselves "as they are," have naturally always been "present." Our understanding of "tree" has therefore changed completely because a tree is now something which is composed of several elements. In that sense, we have changed the original phenomenon by adding something – and this has all happened in our thinking and not in the tree itself. It is also possible to find different structures and different components for the tree, e.g., the brown and the green, which shows that we construct this knowledge.

Next, we can investigate the components to see how they interact with and relate to each other and to the whole system. Also, we can work out their functions and determine the conditions under which a certain event will occur. We can even expand the scope of our investigation and examine the tree in the context of its ecosystem. But no matter what we do or how sophisticated our investigation becomes, everything said above remains true here, too, because neither all these actions listed above nor the knowledge we gain from them are the object itself. Rather, we have added something to the object and the more we know about our object, the more knowledge we have constructed. This addition is what science—gaining knowledge— is all about. Or in the words of Roland Barthes: "the simulacrum is intellect added to object, and this addition has an anthropological value, in that it is man himself, his history, his situation, his freedom and the very resistance which nature offers to his mind" (1963/1972, p. 214/215).

In principle, this holds truth regarding all scientific investigations. But the more complex phenomena are, the more effort and personal contribution is required on behalf of the investigator to come up with structures, theories, or explanations. Paraphrasing Barthes: When dealing with complex phenomena, more intellect must be added to the object, which means in turn that there are more possibilities for different approaches and perspectives, that is, the constructive element becomes larger. As discussed previously, this does not mean that investigative and interpretative processes are arbitrary. But it is clear from this train of thought that "objectivity" or "truth" in a "positivist," naïve empiricist "realist," or absolute sense are not attainable. Nevertheless, we argue here that this is not a drawback, as many critics of postmodernism contend (see above), but rather an advantage because it allows more accurate scientific investigations of true-to-life phenomena, which are typically complex in the case of psychology.

The concepts of simulacra by Baudrillard and Barthes can be combined to provide a description of the experiment in psychology. Accordingly, our understanding of the concept of the "simulacrum" entails that scientific processes – indeed all investigative processes – necessarily need to duplicate the object of their investigation in order to understand it. In doing so, constructive elements are necessarily introduced. These elements are of a varying nature, which means that investigations of one and the same phenomenon may differ from each other and different investigations may find out different things about the phenomenon in question. These investigations then become entities on their own – in the Baudrillardian sense – and therefore simulacra.

In a groundbreaking article on "the meaning and limits of exact science" physicist Max Planck stated that "[a]n experiment is a question which science poses to nature, and a measurement is the recording of nature's answer" (Planck, 1949, p. 325). The act of "asking a question" implies that the person asking the question has at least a general idea of what the answer might look like (Heidegger, 1953, §2). For example: When asking someone for their name, we obviously do not know what they are called, but we assume that they have a name and we also have an idea of how the concept "name" works. Otherwise we could not even conceive, let alone formulate, and pose our question. This highlights how a certain degree of knowledge and understanding of a concept is necessary so that we are able to ask questions about it. Likewise, we need to have a principal idea or assumption of possible mechanisms if we want to find out how more complex phenomena function. It – at least at the beginning –

irrelevant whether these ideas are factually correct or entirely wrong, for without them we would be unable to approach our subject matter in the first place.

The context of the investigator – their general worldview, their previous knowledge and understanding, and their social situation – obviously plays an important part in the process of forming a question which can be asked in the current research context. Although this context may be analyzed along postmodern lines in order to find out how it affects research, production of knowledge, and – when the knowledge is applied – possible (social) consequences, there is a much more profound implication pertaining to the very nature of the experiment as a means to gain knowledge.

Irrespective of whether it is a simple experiment in physics such as Galileo Galilei's or an experiment on a complex phenomenon from social or cognitive psychology, the experiment is a situation which is specifically designed to answer a certain type of questions, usually causal relationships, such as: "Does A causally affect B?" Excluding the extremely complex discussion on the nature of causality and causation (e.g., Armstrong, 1997; Pearl, 2009; Paul & Hall, 2013), it is crucial to note that we need the experiment as a tool to answer this question. Although we may theorize about a phenomenon and infer causal relationships simply by observing, we cannot – at least according to the prevailing understanding of causality in the sciences – prove causal relationships without the experiment.

The basic idea of the experiment is to create conditions which differ in only one single factor which is suspected as a causal factor for an effect. The influence of all other potential causal relationships is kept identical because they are considered as confounding factors which are irrelevant from the perspective of the research question of the current experiment. Then, if a difference is found in the outcome between the experimental conditions, this is considered as proof that the aspect in question exerts indeed a causal effect. This procedure and the logic behind it are not difficult to understand. However, a closer look reveals that this is actually far from simple or obvious.

To begin with, an experiment is nothing which occurs "naturally" but a situation created for a specific purpose, i.e., an "artificial" situation, because other causal factors exerting influence in "real" life outside the laboratory are deliberately excluded and considered as "confounding" factors. This in itself shows that the experiment contains a substantial postmodern element because instead of creating something it rather re-creates it. This re-creation is of course based on phenomena from the "profound" reality – in the Baudrillardian sense – since the explicit aim is to find out something about this profound reality and not to create something new or something else. However, as stated above, this re-creation must contain constructive elements reflecting the presuppositions, conceptual-theoretical assumptions, and aims of the investigator. By focusing on one factor and by reducing the complexity of the profound reality, the practical operationalization and realization thus reflect both the underlying conceptual structure and the anticipated outcome as they are specifically designed to test for the suspected but hidden or obscured causal relationships.

At this point, another element becomes relevant, namely the all-important role of language, which is emphasized in postmodern thinking (e.g., Harris, 2005). Without going into the intricacies of semiotics, there is an explanatory gap (Chalmers, 2005) – to borrow a phrase from philosophy of mind – between the phenomenon on the one hand and the linguistic and/or mental representation of it on the other. This relationship is far from clear and it is therefore problematic to assume that our linguistic or mental representations – our words and the concepts they designate – are identical with the phenomena themselves. Although we cannot, at least according to our present knowledge and understanding, fully bridge this gap, it is essential to be aware of it in order to avoid some pitfalls, as will be shown in the examples below.

Even a seemingly simple word like "tree" – to take up once more our previous example – refers to a tangible phenomenon because there are trees "out there." However, they come in all shapes and sizes, there are different kinds of trees, and every single one of them may be labeled as "tree." Furthermore, trees are composed of different parts, and the leaf – although part of the tree – has its own word, i.e., linguistic and mental representation. Although the leaf is part of the tree – at least according to our concepts – it is unclear whether "tree" also somehow encompasses "leaf." The same holds true for the molecular, atomic, or even subatomic levels, where there "is" no tree. Excluding the extremely complex ontological implications of this problem, it has become clear that we are referring to a certain level of granularity when using the word "tree." The level of granularity reflects the context, aims, and concepts of the investigator, e.g., an investigation of the rain forest as an ecosystem will ignore the subatomic level.

How does this concern experimental psychology? Psychology studies intangible phenomena, namely mental and behavioral processes, such as cognition, memory, learning, motivation, emotion, perception, consciousness, etc. It is important to note that these terms designate theoretical constructs as, for example, memory cannot be observed directly. We may provide the subjects of an experiment a set of words to learn and observe later how many words they reproduce correctly. A theoretical construct therefore describes such relationships between stimulus and behavior, and we may draw conclusions from this observable data about memory. But neither the observable behavior of the subject, the resulting data, nor our conclusions are identical with memory itself.

This train of thought demonstrates the postmodern character of experimental psychology because we construct our knowledge. But there is more to it than that: Even by trying to define a theoretical construct as exactly as possible – e.g., memory as "the process of maintaining information over time" (Matlin, 2012, p. 505) or "the means by which we retain and draw on our past experiences to use this information in the present" (Sternberg & Sternberg, 2011, p. 187) – the explanatory gap between representation and phenomenon cannot be bridged. Rather, it becomes even more complicated because theoretical constructs are composed of other theoretical constructs, which results in some kind of self-referential circularity where constructs are defined by other constructs which refer to further constructs. In the definitions above, for instance, hardly any key term is self-evident and unambiguous for there are different interpretations of the constructs "process," "maintaining," "information," "means," "retain," "draw on," "experiences," and "use" according to their respective contexts. Only the temporal expressions "over time," "past," and "present" are probably less ambiguous here because they are employed as non-technical, everyday terms. However, the definitions above are certainly not entirely incomprehensible – in fact, they are rather easy to understand in everyday language – and it is quite clear what the authors intend to express. The italics indicate constructive elements, which demonstrates that attempts to give a precise definition in the language of science result in fuzziness and self-reference.

Based on a story by Jorge Luis Borges, Baudrillard (1981) found an illustrative allegory: a map so precise that it portrays everything in perfect detail – but therefore inevitably so large that it shrouds the entire territory it depicts. Similarly, Taleb (2007) coined the term "ludic fallacy" for mistaking the model/map – in our context: experiments in psychology – for the reality/territory, that is, a mental or behavioral phenomenon. Similar to the functionality of a seemingly "imprecise" map which contains only the relevant landmarks so the user may find their way, the fuzziness of language poses no problems in everyday communication. So why is it a problem in experimental psychology? Since the nature of theoretical constructs in psychology lies precisely in their very fuzziness, the aim of reaching a high degree of granularity and precision in experimental psychology seems to be unattainable (see the various failed attempts to create "perfect" languages which might depict literally everything "perfectly," e.g., Carapezza & D'Agostino, 2010).

Without speculating about ontic or epistemic implications, it is necessary to be aware of the explanatory gap and to refrain from identifying the experiment and the underlying operationalization with the theoretical construct. Otherwise, this gap is "filled" unintentionally and uncontrollably if the results of an experiment are taken as valid proof for a certain theoretical construct, which is actually fuzzy and potentially operationalizable in a variety of ways. If this is not acknowledged, words, such as "memory," become merely symbols devoid of concrete meaning, much like a glass bead game – or in postmodern terminology: a hyperreality.

Experiments and Hyperreality

"Hyperreality" is another key term in the work of Jean Baudrillard (1981) and it denotes a concept closely related to the simulacrum. Accordingly, in modern society the simulacra are ubiquitous and they form a system of interconnected simulacra which refer to each other rather than to the real, thereby possibly hiding or replacing the real. Consequently, the simulacra become real in their own right and form a "more real" reality, namely the hyperreality. One may or may not accept Baudrillard's conception, especially the all-embracing social and societal implications, but the core concept of "hyperreality" is nevertheless a fruitful tool to analyze experimental psychology. We have already seen that the experiment displays many characteristics of a simulacrum, so it is not surprising that the concept of hyperreality is applicable here as well, although in a slightly different interpretation than Baudrillard's.

The hyperreal character of the experiment can be discussed on two levels: the experiment itself and the discourse wherein it is embedded.

On the level of the experiment itself, two curious observations must be taken into account. First, and in contrast to the natural sciences where the investigator is human and the subject matter (mostly) non-human and usually inanimate, in psychology both the investigator and the subject matter are human. This means that the subjects of the experiment, being autonomous persons, are not malleable or completely controllable by the investigator because they bring their own background, history, worldview, expectations, and motivations. They interpret the situation – the experiment – and act accordingly, but not necessarily in the way the investigator had planned or anticipated (Smedslund, 2016). Therefore, the subjects create their own versions of the experiment, or, in postmodern terminology, a variety of simulacra, which may be more or less compatible with the framework of the investigator. This holds true for all subjects of an experiment, which means that the experiment as a whole may also be interpreted as an aggregation of interconnected simulacra – a hyperreality.

The hyperreal character becomes even more evident because what contributes in the end to the interpretation of the results of the experiment are not the actual performances and results of the individual subjects as they were intended by them but rather how their performances and results are handled, seen, and interpreted by the investigator. Even if the investigator tries to be as faithful as possible and aims at an exact and unbiased measurement – i.e., an exact copy – there are inevitably constructive elements which introduce uncertainty into the experiment. Investigators can never be certain what the subjects were actually doing and thinking so they must necessarily work with interpretations. Or in postmodern terms: Because the actual performances and results of the subjects are not directly available the investigators must deal with simulacra. These simulacra become the investigators' reality and thus any further treatment – statistical analyses, interpretations, or discussions – becomes a hyperreality, that is, a set of interconnected simulacra which have become "real."

On the level of the discourse wherein the experiment is embedded, another curious aspect also demonstrates the hyperreal character of experimental psychology. Psychology is, according to the standard definition, the scientific study of mental and behavioral processes of the individual (e.g., Gerrig, 2012). This definition contains two actually contradictory elements. On the one hand, the focus is on processes of the individual. On the other hand, the – scientific – method to elucidate these processes does not look at individuals per se but aggregates their individual experiences and transforms them into a "standard" experience. The results from experiments, our knowledge of the human psyche, reflect psychological functioning at the level of the mean across individuals. And even if we assume that the mean is only an estimator and not an exact description or prediction, the question remains open how de-individualized observations are related to the experience of an individual. A general mechanism, a law – which was discovered by abstracting from a multitude of individual experiences – is then (re-)imposed in the opposite direction back onto the individual. In other words, a simulacrum – namely, the result of an experiment – is viewed and treated as reality, thus becoming hyperreal. Additionally, and simply because it is considered universally true, this postulated law acquires thereby a certain validity and "truth" – often irrespective of its actual, factual, or "profound" truth – on its own. Therefore, it can become impossible to distinguish between "profound" and "simulacral" truth, which is the hallmark of hyperreality.

Measuring the Capacity of the Visual Working Memory

Vision is an important sensory modality and there is extensive research on this area (Hutmacher, 2019b). Much of our daily experience is shaped by seeing a rich and complex world around us, and it is therefore an interesting question how much visual information we can store and process. Based on the development of a seminal experimental paradigm, Luck and Vogel (1997) have shown that visual working memory has a storage capacity of about four items. This finding is reported in many textbooks (e.g., Baddeley, 2007; Parkin, 2013; Goldstein, 2015) and has almost become a truism in cognitive psychology.

The experimental paradigm developed by Luck and Vogel (1997) is a prime example of an experiment which closely adheres to the scientific principles outlined above. In order to make a very broad and fuzzy phenomenon measurable, simple abstract forms are employed as visual stimuli – such as colored squares, triangles, or lines, usually on a "neutral," e.g., gray, background – which can be counted in order to measure the capacity of visual working memory. Reducing the exuberant diversity of the "outside visual world" to a few abstract geometric forms is an extremely artificial situation. The obvious contrast between simple geometrical forms and the rich panorama of the "real" visual world illustrates the pitfalls of controlling supposed confounding variables, namely the incontrollable variety of the "real" world and how we see it. Precisely by abstracting and by excluding potential confounding variables it is possible to count the items and to make the capacity of the visual working memory measurable. But in doing so the original phenomenon – seeing the whole world – is lost. In other words: A simulacrum has been created.

The establishment of the experimental paradigm by Luck and Vogel has led to much research and sparked an extensive discussion how the limitation to only four items might be explained (see the summaries by Brady et al., 2011; Luck & Vogel, 2013; Ma et al., 2014; Schurgin, 2018). However, critically, several studies have shown that the situation is different when real-world objects are used as visual stimuli rather than simple abstract forms, revealing that the capacity of the visual working memory is higher for real-world objects (Endress & Potter, 2014; Brady et al., 2016; Schurgin et al., 2018; Robinson et al., 2020; also Schurgin & Brady, 2019). Such findings show that the discourse about the mechanisms behind the limitations of the visual working memory is mostly about an artificial phenomenon which has no counterpart in "reality" – the perfect example of a hyperreality.

This hyperreal character does not mean that the findings of Luck and Vogel (1997) or similar experiments employing artificial stimuli are irrelevant or not "true." The results are true

– but it is a local truth, only valid for the specific context of specific experiments, and not a global truth which applies to the visual working memory in general. That is, speaking about "visual working memory" based on the paradigm of Luck and Vogel is a mistake because it is actually about "visual working memory for simple abstract geometrical forms in front of a gray background."

Free Will and Experimental Psychology

The term "free will" expresses the idea of having "a significant kind of control [italics in the original] over one's actions" (O'Connor & Franklin, 2018, n.p.). This concept has occupied a central position in Western philosophy since antiquity because it has far-reaching consequences for our self-conception as humans and our position in the world, including questions of morality, responsibility, and the nature of legal systems (e.g., Beebee, 2013; McKenna & Pereboom, 2016; O'Connor & Franklin, 2018). Being a topic of general interest, it is not surprising that experimental psychologists have tried to investigate free will as well.

The most famous study was conducted by Libet et al. (1983), and this experiment has quickly become a focal point in the extensive discourse on free will because it provides empirical data and a scientific investigation. Libet et al.'s experiment seems to show that the subjective impression when persons consciously decide to act is in fact preceded by objectively measurable but unconscious physical processes. This purportedly proves that our seemingly voluntary actions are actually predetermined by physical processes because the brain has unconsciously reached a decision already before the person becomes aware of it and that our conscious intentions are simply grafted onto it. Therefore, we do not have a free will, and consequently much of our social fabric is based on an illusion. Or so the story goes.

This description, although phrased somewhat pointedly, represents a typical line of thought in the discourse on free will (e.g., the prominent psychologists Gazzaniga, 2011; Wegner, 2017; see Kihlstrom, 2017, for further examples).

Libet's experiment sparked an extensive and highly controversial discussion: For some authors, it is a refutation or at least threat to various concepts of free will, or, conversely, an indicator or even proof for some kind of material determinism. By contrast, other authors deny that the experiment refutes or counts against free will. Furthermore, a third group – whose position we adopt for our further argumentation – denies that Libet's findings are even relevant for this question at all (for summaries of this complex and extensive discussion and various positions including further references see Nahmias, 2010; Radder & Meynen, 2013; Schlosser, 2014; Fischborn, 2016; Lavazza, 2016; Schurger, 2017). Libet's own position, although not entirely consistent, opposes most notions of free will (Roskies, 2011; Seifert, 2011). Given this background, it is not surprising that there are also numerous further experimental studies on various aspects of this subject area (see the summaries by Saigle et al., 2018; Shepard, 2018; Brass et al., 2019).

However, we argue that this entire discourse is best understood along postmodern lines as hyperreality and that Libet's experiment itself is a perfect example of a simulacrum. A closer look at the concrete procedure of the experiment shows that Libet actually asked his participants to move their hand or finger "at will" while their brain activity was monitored with an EEG. They were instructed to keep watch in an introspective manner for the moment when they felt the "urge" to move their hand and to record this moment by indicating the clock-position of a pointer. This is obviously a highly artificial situation where the broad and fuzzy concept of "free will" is abstracted and reduced to the movement of the finger, the only degree of freedom being the moment of the movement. The question whether this is an adequate operationalization of free will is of paramount importance, and there are many objections that Libet's setup fails to measure free will at all (e.g., Mele, 2007; Roskies, 2011; Kihlstrom, 2017; Brass et al., 2019).

Before Libet, there was no indication that the decision when to move a finger might be relevant for the concept of free will and the associated discourse. The question whether we have control over our actions referred to completely different levels of granularity. Free will was discussed with respect to questions such as whether we are free to live our lives according to our wishes or whether we are responsible for our actions in social contexts (e.g., Beebee, 2013; McKenna & Pereboom, 2016; O'Connor & Franklin, 2018), and not whether we lift a finger now or two seconds later. Libet's and others' jumping from very specific situations to far-reaching conclusions about a very broad and fuzzy theoretical construct illustrates that an extremely wide chasm between two phenomena, namely moving the finger and free will, is bridged in one fell swoop.

In other words, Libet's experiment is a simulacrum as it duplicates a phenomenon from our day-to-day experience – namely free will – but in doing so the operationalization alters and reduces the theoretical construct. The outcome is a questionable procedure whose relationship to the phenomenon is highly controversial. Furthermore, the fact that, despite its tenuous connection to free will, Libet's experiment sparked an extensive discussion on this subject reveals the hyperreal nature of the entire discourse because what is being discussed is not the actual question – namely free will – but rather a simulacrum. Everything else – the arguments, counter-arguments, follow-up experiments, and their interpretations – built upon Libet's experiment are basically commentaries to a simulacrum and not on the real phenomena. Therefore, a hyperreality is created where the discourse revolves around entirely artificial phenomena, but where the arguments in this discussion refer back to and affect the real as suggestions are made to alter the legal system and our ideas of responsibility – which, incidentally, is not a question of empirical science but of law, ethics, and philosophy.

All of the above is not meant to say that this whole discourse is meaningless or even gratuitous – on the contrary, our understanding of the subject matter has greatly increased. Although our knowledge of free will has hardly increased, we have gained much insight into the hermeneutics and methodology – and pitfalls! – of investigations of free will, possible consequences on the individual and societal level, and the workings of scientific discourses. And this is exactly what postmodernism is about.

Discussion

As shown above, there are a number of postmodern elements in the practice of experimental psychology: The prominent role of language, the gap between the linguistic or mental representation and the phenomenon, the "addition of intellect to the object," the simulacral character of the experiment itself in its attempt to re-create phenomena, which necessarily transforms the "real" phenomenon due to the requirements of the experiment, and finally the creation of a hyperreality if experiments are taken as the "real" phenomenon and the scientific discourse becomes an exchange of symbolic expressions referring to the simulacra created in experiments, replacing the real. All these aspects did not seep gradually into experimental psychology in the wake of postmodernism but have been present since the very inception of experimental psychology as they are necessarily inherent in its philosophy of science.

Given these inherent postmodern traits in experimental psychology, it is puzzling that there is so much resistance against a perceived "threat" of psychology's scientificness. Although a detailed investigation of the reasons lies outside the scope of this analysis, we suspect there are two main causes: First, an insufficient knowledge of the history of science and understanding of philosophy of science may result in idealized concepts of a "pure" natural science. Second, lacking familiarity with basic tenets of postmodern approaches may lead to the assumption that postmodernism is just an idle game of arbitrary words. However, "science" and “postmodernism” and their respective epistemological concepts are not opposites (Gergen, 2001; Holtz, 2020). This is especially true for psychology, which necessarily contains a social dimension because not only the investigators are humans but also the very subject matter itself.

The (over-)reliance on quantitative-experimental methods in psychology, often paired with a superficial understanding of the philosophy of science behind it, has been criticized, either from the theoretical point of view (e.g., Bergmann & Spence, 1941; Hearnshaw, 1941; Petrie, 1971; Law, 2004; Smedslund, 2016) or because the experimental approach has failed to produce reliable, valid, and relevant applicable knowledge in educational psychology (Slavin, 2002). It is perhaps symptomatic that a textbook teaching the principles of science for psychologists does not contain even one example from experimental psychology but employs only examples from physics, plus Darwin’s theory of evolution (Wilton & Harley, 2017).

On the other hand, the postmodern perspective on experimental psychology provides insight into some pitfalls, as illustrated by the examples above. On the level of the experiment, the methodological requirements imply the creation of an artificial situation, which opens up a gap between the phenomenon as it is in reality and as it is concretely operationalized in the experimental situation. This is not a problem per se as long as is it clear – and clearly communicated! – that the results of the experiment are only valid in a certain context. The problems begin if the movement of a finger is mistaken for free will. Similarly, being aware that local causalities do not explain complex phenomena such as mental and behavioral processes in their entirety also prevents (over-)generalization, especially if communicated appropriately. These limitations make it clear that the experiment should not be made into an absolute or seen as the only valid way of understanding the psyche and the world.

On the level of psychology as an academic discipline, any investigation must select the appropriate level of granularity and strike a balance between the methodological requirements and the general meaning of the theoretical concept in question to find out something about the “real” world. If the level of granularity is so fine that results cannot be tied back to broader theoretical constructs rather than providing a helpful understanding of our psychological functioning, academic psychology is in danger of becoming a self-referential hyperreality.

The postmodern character of experimental psychology also allows for a different view on the so-called replication crisis in psychology. Authors contending that there is no replication crisis often employ arguments which exhibit postmodern elements, such as the emphasis on specific local conditions in experiments which may explain different outcomes of replication studies (Stroebe & Strack, 2014; Baumeister, 2019). In other words, they invoke the simulacral character of experiments. This explanation may be valid or not, but the replication crisis has shown the limits of a predominantly experimental approach in psychology.

Acknowledging the postmodern nature of experimental psychology and incorporating postmodern thinking explicitly into our research may offer a way out of this situation. Our subject matter—the psyche—is extremely complex, ambiguous, and often contradictory. And postmodern thinking has proven capable of successfully explaining such phenomena (e.g., Bertens, 1995; Sim, 2011; Aylesworth, 2015). Thus, paradoxically, by accepting and considering the inherently fuzzy nature of theoretical constructs, they often become much clearer (Ronzitti, 2011). Therefore, thinking more along postmodern lines in psychology would actually sharpen the theoretical and conceptual basis of experimental psychology – all the more as experimental psychology has inevitably been a postmodern endeavor since its very beginning.

2.3 Study 3 – Re-examining the testing effect as a learning strategy: the advantage of retrieval practice over concept mapping as a methodological artifact

This is a pre-copy-editing, author-produced version of an article published 2023 in Frontiers in Psychology (Cognitive Science) following peer review. The official citation that should be used in referencing this material is: Mayrhofer, R., Kuhbandner, C., & Frischholz, K. (2023). Re-examining the testing effect as a learning strategy: the advantage of retrieval practice over concept mapping as a methodological artifact. *Frontiers in Psychology, 14*, 1258359. https://doi.org/10.3389/fpsyg.2023.1258359.

Abstract

Several previous studies appear to have demonstrated that studying with retrieval practice produces more learning than studying with concept mapping, a finding based on which an extended use of retrieval practice in educational practice was recommended. However, a closer examination of the methods used in these previous studies reveals a crucial confounding variable: Whereas participants in the concept mapping conditions performed a concept mapping task without any subsequent memorizing of the learning material, participants in the retrieval practice conditions performed not only retrieval practice but also an additional memorization task, which doubled the total memorization time. The present preregistered study examined whether the advantage observed in the retrieval practice condition over the concept mapping condition in previous studies was actually driven by additional memorization rather than by retrieval practice. While we replicated the previous finding that retrieval practice in combination with additional memorizing produces more learning than concept mapping without additional memorizing, this advantage of retrieval practice over concept mapping vanished when participants in the concept mapping condition, too, memorized the learning material after having created a concept map. These findings demonstrate that the assumed advantage of retrieval practice over concept mapping in fact represents a methodological artifact. Besides serving as a reminder of the importance of a solid methodology, the present study also illustrates the importance of using of an adequate terminology. Depicting a learning strategy condition as "retrieval practice" when the condition actually encompasses not only retrieval practice but also additional memorizing obfuscates the possibility that observed memory advantages may not be fueled by retrieval practice, i.e., the learning strategy as such. We conclude by giving an outlook on the ramifications of our findings for cognitive and educational psychology.

Keywords: testing effect, retrieval practice, concept mapping, learning strategies, educational psychology, cognitive psychology

Introduction

The effectiveness of learning strategies is an important topic that is extensively researched in applied research. In a highly prominent and frequently cited study, Karpicke and Blunt (2011) investigated in the context of learning strategies an important finding from basic research, namely, the so-called testing effect, which describes the phenomenon that retrieval enhances long-term memory. They came to the conclusion that retrieval practice produces more learning than elaborative studying using concept mapping. A virtually identical result was found by O'Day and Karpicke (2021), who employed the same methodology as Karpicke and Blunt (2011). These results were also found and therefore confirmed by Lechuga et al. (2015) and Camerer et al. (2018), who also employed the same methodology as Karpicke and Blunt (2011), although the advantage of retrieval practice was notably smaller than compared to the work of Karpicke and Blunt (2011). In the light of the far-reaching ramifications for both cognitive and educational psychology if in fact retrieval practice really does produce more and better learning than elaborative studying with concept mapping, it is evidently important to ascertain that the basis for such propositions is theoretically and methodologically solid. This is why this study re-examined and empirically tested the proposition that retrieval practice produces more learning than elaborative studying with concept mapping, focusing primarily on the methodology of previous experiments.

Karpicke and Blunt (2011) as well as O'Day and Karpicke (2021) conclude that retrieval practice is a better learning strategy because they report to have empirically shown that retrieval practice produces more learning than elaborative studying with concept mapping. Specifically, their conclusion is based on their finding that performance in a memory test was better in a retrieval practice condition compared to a concept mapping condition. We propose, however, that the reasons for the better performance in the retrieval practice condition, as found by Karpicke and Blunt (2011) and O'Day and Karpicke (2021), and, by extension, also in the studies by Lechuga et al. (2015) and Camerer et al. (2018), which employ the same methodology, are not based on certain specific cognitive mechanisms inherent to or associated with the respective learning strategy. Instead, there is reason to assume that the better performance in the retrieval practice condition occurred due to a methodological artifact because a closer analysis of the methods employed by Karpicke and Blunt (2011) reveals two potential confounders inherent in the design and execution of these studies, which might have biased the observed results. These potential confounders also affect the studies by Lechuga et al. (2015), Camerer et al. (2018), and O'Day and Karpicke (2021).

The first potential confounder pertains to Karpicke and Blunt's (2011) operationalization of what they refer to as "retrieval practice." At the beginning of their experiment, participants in all conditions were asked to study a text about sea otters for a 1-week delayed memory test for 5 min. After that, conditions differed, but a closer look reveals that the conditions differed not only – as the designation as "retrieval practice" and "concept mapping" suggests – in that the participants performed retrieval practice in one condition and concept mapping in the other. Rather, in the "retrieval practice" condition, participants performed a memorization task in addition to retrieval practice, as illustrated below in Figure 1. In this memorization task, they were asked to memorize the text for 5 min. By contrast, in the concept mapping condition, participants only created their concept map, and there was no additional memorization task. In particular, participants in the concept mapping condition were instructed not to invest any additional time in memorizing the material as they were told "that if they finished before the end of the 25-min period, they should spend the remaining amount of time reviewing their map and making sure they had included all the details from the text in their map" (Karpicke & Blunt, 2011; Supporting Online Material, p. 2).

This difference between conditions, when there is an additional memorization phase in one condition but not in the other, is problematic for at least three reasons. First, from a methodological perspective, it seems likely that the advantage observed by Karpicke and Blunt (2011) in the retrieval practice condition over the concept mapping condition was not actually driven by retrieval practice but rather by the additional memorization phase, which doubled the time the participants had in the retrieval practice condition to memorize the learning material for the later test.

There is substantial evidence reaching back as early as Ebbinghaus (1885/1913) that learning performance increases with increased memorization time (Murdock, 1960; Bugelski, 1962; Zacks, 1969; Geiselman, 1977; Fredrick & Walberg, 1980; Gettinger, 1991; Cook et al., 2010; Yang et al., 2016; Chen & Yang, 2020). This indicates, roughly speaking, that learning performance increases with increased memorization time.

Furthermore, the difference in memorization time between the conditions might also act as confounder in another way. According to the well-established spacing effect (e.g., Rohrer & Pashler, 2007; De Jonge et al., 2012; Kim et al., 2019; Murphy et al., 2022), distributed learning is more effective than massed learning. Therefore, considering that in the retrieval practice condition in Karpicke and Blunt's (2011) study participants memorized the learning material in two study phases at different time points during the experiment (in the initial study phase and the subsequent retrieval practice phase, see Figure 1), this condition represents an example for distributed learning. By contrast, in the concept mapping condition, participants memorized the learning material in only one study phase (in the initial study phase), which represents massed learning. This could indicate that the spacing effect may additionally have contributed to the observed advantage in the retrieval practice condition. This is further supported by a demonstration that spacing also affects the testing effect, as reported by Carpenter and DeLosh (2005), who found that the effect of testing increases with spaced learning.

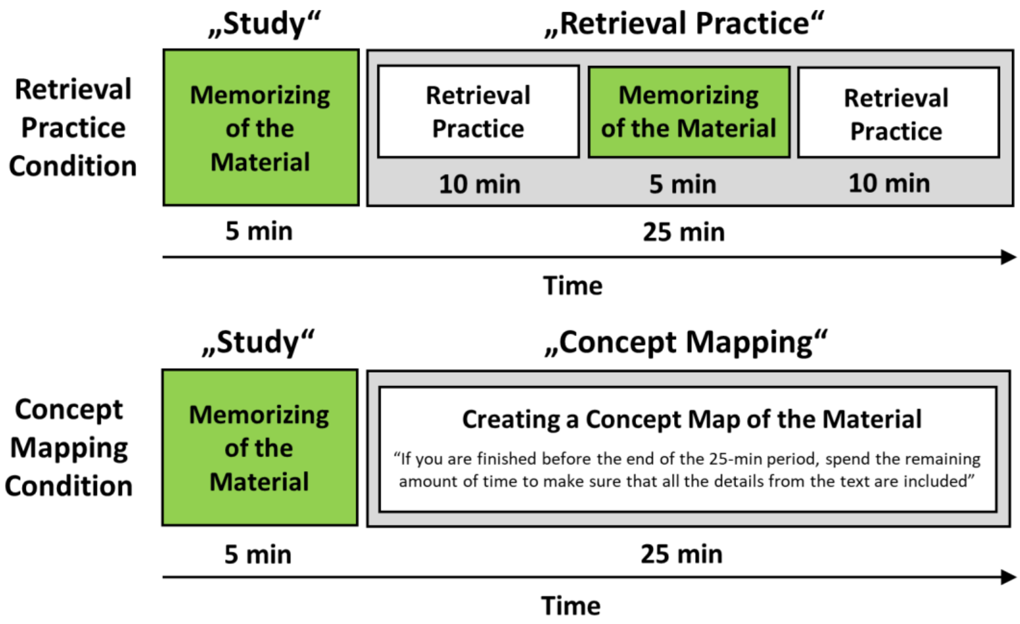

***Figure 1*. Illustration of the confounding variable "Memorization" in Karpicke and Blunt's (2011) study.** The terms used to describe the different learning strategy conditions, namely, "retrieval practice" and "concept mapping," give the impression that only retrieval practice or concept mapping, respectively, were performed in each condition. However, in the so-called "retrieval practice" condition, participants not only performed a retrieval task but also an additional memorization task. By contrast, in the concept mapping condition, participants only performed a concept mapping task without any additional memorization of the learning material. The additional memorization task doubled the time participants spent memorizing the learning material for the later test in the retrieval practice condition. Note that the text the participants were to learn was available during the creation of the concept map but not during retrieval practice.

Second, from a theoretical perspective, the conceptual terms used by Karpicke and Blunt (2011) seem inaccurate. In the title and throughout the whole paper, they state that retrieval practice is a better learning strategy than concept mapping. However, this terminology is inaccurate as their so-called "retrieval practice" condition actually encompasses not only retrieval practice but also an additional memorization phase. Thus, "retrieval practice" is actually operationalized by a combination of two learning strategies, namely, retrieval practice and memorizing. Therefore, the correct conclusion from Karpicke and Blunt's (2011) study should be that retrieval practice in combination with additional memorization produces more learning than concept mapping without additional memorization, which accurately reflects their actual operationalization.

Third, from an applied perspective, it seems doubtful that Karpicke and Blunt's (2011) results can be transferred beyond the laboratory and applied to real-life learning contexts. When preparing for a test where the ability to retrieve memorized facts is measured, it seems unlikely that learning is done as the participants did in Karpicke and Blunt's (2011) concept mapping condition. The purpose of concept mapping is to structure and organize the content of material that should be learned in order to facilitate its understanding (Novak, 1995; Novak & Cañas, 2006) but not to commit this material to memory for a later memory test. In order to achieve the latter goal, additional memorization strategies beyond establishing a conceptual structure of the text must be used. This is the reason why, according to established text learning techniques such as PQ4R (Thomas & Robinson, 1972), additional activities must follow in order to commit the content to memory so that the content can be successfully retrieved later.

Summing up, the fact that in Karpicke and Blunt's (2011) study there was an additional memorization phase in the retrieval practice condition but not in the concept mapping condition is problematic from methodological, theoretical, and applied perspectives.

The second potential confounder pertains to the instructions given in Karpicke and Blunt's (2011) experiment. Here, there is also a critical difference between conditions. In the retrieval practice task, the following instruction was provided above the box where the recalled information had to be typed in: "Please use the space in the box below to write as much information as you can recall about the Sea Otters text you just read" (personal communication with J.R. Blunt). Thus, while performing the retrieval practice task, participants were explicitly prompted that the task was to retrieve and memorize literally everything from the text.

By contrast, in the concept mapping task, the following instruction was provided above the box where the concept map had to be created: "Please use the space below to create your concept map about the Sea Otters text" (personal communication with J.R. Blunt). Only in the instruction provided before it was mentioned "that if they finished [the concept map] before the end of the 25-min period, they should spend the remaining amount of time reviewing their map and making sure they had included all the details from the text in their map" (Karpicke & Blunt, 2011, Supporting Online Material, p. 2). That is, while performing the concept mapping task, other than in the retrieval practice condition, the participants were not prompted that all information from the text should be included in the created concept map.

Using different instructions, which in one condition but not in the other emphasize that the text should be stored in a way that as much information as possible can be retrieved, may have contributed to the observed difference in the final test performance between the retrieval practice condition and the concept mapping condition. Previous research has shown that the instruction to focus on specific aspects of the learning material while studying can influence the quality of later memory (e.g., McCrudden et al., 2005; Roelle et al., 2015; García-Rodicio, 2023). Therefore, using different instructions in the retrieval practice condition and the concept mapping condition in Karpicke and Blunt (2011) may have led to a different amount of information being processed in the retrieval practice condition vs. the concept mapping condition. Indeed, in Karpicke and Blunt's (2011) study, descriptively, the proportion of idea units recalled in the retrieval task was higher than the proportion of idea units included in the concept maps (0.81 vs. 0.78). However, given their sample size (20 participants per condition), only large effects can be reliably detected, i.e., $d > 0.91$ with 80% probability. Therefore, it is not possible to assess whether this difference reflects a true effect or not.

Furthermore, concept mapping was not designed as a tool to study as many details of a text as possible but rather as a tool to structure and organize knowledge (Novak, 1995; Novak & Cañas, 2006). Considering that the participants in Karpicke and Blunt's (2011) concept mapping condition were "instructed about the nature of concept mapping [and] viewed an example of a concept map" (p. 773), it seems likely that the participants viewed concept mapping as a tool to build a mental structure of the relevant contents of a text rather than a tool to foster the ability to later retrieve as much information from the text as possible. Since participants were not prompted in the direction of a potential recall of information during the creation of the concept map, participants' focus during the creation of the concept map may have been to build the best possible content structure of the text rather than learning all of the details contained in the text. By contrast, the participants in the retrieval practice condition were – while working on the retrieval practice task – explicitly instructed that they should learn the information and details from the text. Since test performance in the final test was mainly determined by the ability to remember as many details from the text as possible, the difference in focus during learning may thus have contributed to the observed advantage of the retrieval practice over the context mapping condition.

In summary, there are two potential confounders in the paradigm used by Karpicke and Blunt (2011) which favor the retrieval practice condition over the concept mapping condition and may thus offer an alternative explanation for the observed performance advantage of the retrieval practice condition over the concept mapping condition. The aim of the present study was to re-examine this issue and to rule out that the reported advantage of retrieval practice over concept mapping in previous studies may actually stem from unnoticed confounders.

To this end, we conducted an experiment which was specifically designed to address the potential confounders as explained above. To avoid the problem of unclear terminology found in previous studies, it is necessary to precisely define the terms used to designate specific cognitive processes. In the present study, "memorizing" is understood as the activity of taking in and storing learning material with the aim of retaining it over a longer period of time in order to be able to recall and reproduce it later. "Retrieval practice" means that participants retrieve previously studied material from memory. "Concept mapping" is understood as the activity of structuring and organizing the content of the learning material in form of a concept map.

Besides the exact replication of Karpicke and Blunt's (2011) original retrieval practice and concept mapping conditions, two additional concept mapping conditions were added (see Figure 2 below). In one condition, to control for the additional memorization in the retrieval practice condition, participants in the concept mapping condition were tasked to memorize the concept map they created, i.e., memorization time in this condition was as long as in the retrieval practice condition, namely 10 min. In the other condition, to control for differences in instructions, participants were instructed during the concept mapping task to create a concept map that contains as many details of the text as possible.

We expected to replicate the findings reported by Karpicke and Blunt (2011), that is, we expected that performance in the final test would be higher in the original retrieval practice condition (with additional memorization) than in the original concept mapping condition (without additional memorization). If the advantage of the retrieval practice condition over the concept mapping condition is actually driven by the additional memorization in the original retrieval practice condition, the advantage of the retrieval practice condition should decrease or even disappear if a second memorization period—after the creation of the concept map—is present. If the advantage of the retrieval practice condition over the concept mapping condition is actually driven by the differences in the instructions used in the original conditions, the advantage of the retrieval practice condition should decrease or even disappear if participants are prompted during the creation of the concept map as well that the concept map should contain as many details of the text as possible.

Materials and Methods

All materials, procedures and statistical tests followed our preregistration at Open Science Foundation[1] (see https://osf.io/zx7h2). According to German law, no ethics approval was required as there were no potential negative consequences for the participants of this study.

Participants

A power analysis (G*Power 3.1.9.7; Faul et al., 2007) was used to determine the sample size. Based on a meta-analysis of retrieval practice in the context of teaching by Schwieren et al. (2017), which revealed an overall effect size of $d = 0.56$, the sample size was chosen to be large enough to detect effects of $f = 0.28$ with 95% probability for a one-way ANOVA with four groups ($\alpha = 0.05$). A total of 240 participants were tested[2]; 10 had to be excluded because they were already familiar with the text they were assigned to learn, resulting in a final sample size of $N = 230$. Note that the chosen effect size is more conservative than the effect sizes of $d = 1.50$ found in Karpicke and Blunt (2011) or $d = 0.96$ (verbatim questions) and $d = 0.62$ (inference questions) found by Lechuga et al. (2015), and that the number of participants per condition was about three times higher than in the original study by Karpicke and Blunt (2011).

Materials

Since the present study was a re-examination of Karpicke and Blunt (2011), the very same materials – translated into German – were employed in this study: The learning material consisted of a text of 277 words (275 in the original English text) on the subject of the sea otter. The final test comprised 16 questions: There were 14 verbatim questions, 12 of which yielded 1 scoring point each, 1 question yielded 2 points and 1 question yielded 7 points, totaling 21 scoring points. Furthermore, there were 2 inference questions, each yielding 2 scoring points. Therefore, a maximum of 25 points in total could be achieved. The answers for these questions

[1] As outlined in the preregistration, this experiment further encompassed additional, new inference questions. The results will be addressed in a future article as this article focuses on methodology, based on an exact replication of Karpicke and Blunt (2011). The participants answered those additional inference questions on a separate sheet after completing the original experiment and test from Karpicke and Blunt's (2011) study. Hence, the findings reported here were not influenced by the additional inference questions.

[2] The preregistration stated that the target sample was $N = 160$. This number was increased since statements about the absence of an effect are more reliable when the power is higher, and we chose a sample size that allows the detection of a possible effect with 95% probability, i.e., $N = 224$.

were scored identical to Karpicke and Blunt (2011), meaning that only answers which were considered correct in their experiment were considered correct in our study. All other answers were considered false. All answers were rated by two independent raters, whose mutual agreement was very high: They agreed on 4792 out of 4830 scoring points (99.2%) for the verbatim questions. For the inference questions, the raters agreed in 874 out of 920 (95.0%) scoring points. The remaining 38 and 46 cases were solved by discussion until agreement was reached. The result of the final test is given as percentage of the maximum possible score, i.e., 21 points for the verbatim questions and 4 points for the inference questions.

Procedure

A one-by-four between-subjects design was employed, with learning strategy in combination with potential confounders as factor and the following conditions as factor levels: retrieval practice with additional memorization and with additional instruction "recall as much as possible" (RP + AM + AI condition; original retrieval practice condition as in Karpicke & Blunt, 2011), concept mapping without additional memorization and without additional instruction "incorporate as much as possible" (CM - AM - AI condition; original concept mapping condition as in Karpicke & Blunt, 2011), concept mapping without additional memorization and with additional instruction "incorporate as much as possible" (CM -AM+ AI condition), and concept mapping with additional memorization and with additional instruction "incorporate as much as possible" (CA + AM + AI condition).

The experiment consisted of two sessions conducted in person. In the learning session, participants studied the learning material according to different learning strategies. One week later, in the testing session, participants answered the final test (identical to Karpicke & Blunt, 2011). Participants were tested in groups of up to four persons, although each participant had their own individual, separate cubicle.

At the beginning of the experiment, all participants received general written instructions that they were to learn a text and that they would be tested 1 week later. The instructions stated that all information from the text should be memorized. In all conditions, participants were given the appropriate timeframe of the particular condition (see below). In the three concept mapping conditions, participants were also given a short written instruction, including a graphic example, on the nature of concept maps and how concept maps work. Although Mintzes et al. (2011) criticized Karpicke and Blunt (2011) on the grounds that working with concept maps must be learned thoroughly over a longer period of time and cannot be taught ad hoc by means of a short instruction, our focus here lies on the methodology of Karpicke and Blunt's (2011) experiment. Thus, even if studying with concept mapping is more efficient with more experience (see also Lechuga et al., 2015), the methodology of the experiment would not be affected. Hence, we retained Karpicke and Blunt's (2011) original procedure.

For the learning session, the overall duration of the learning phase was 30 min in all conditions. In all conditions, participants initially had 5 min to study the text (identical to Karpicke & Blunt, 2011). After this point, the conditions differed: In the RP + AM + AI condition, the text was removed in the first recall phase and participants were asked to write down as much as they could recall from the text they just learned. They were given 10 min for this task before they memorized the text once more for a period of 5 min, followed by a second recall phase of 10 min. In the CM - AM - AI condition, participants kept the text for the whole duration of the studying time; participants in this condition then had 25 min to create their concept map on a sheet which simply stated that the concept map should be created below. In the CM - AM + AI condition, the text was also left with the participants for the whole time, who also had 25 min to create their concept map. However, in this condition, the instruction on the sheet for the concept map explicitly stated that the concept map should be created below and that as much information as possible from the text should be incorporated in doing so. This instruction was analogous to the instruction in the retrieval practice condition for the retrieval practice task, which stated that the participants should recall as much information as possible. In the CM + AM + AI condition, the text was also left with the participants, who then had 20 min to create their concept maps. The instruction on the sheet for the concept map stated that the concept map should be created below and that as much information as possible from the text should be incorporated in doing so. After 20 min, the participants were asked to memorize the concept maps they had just created for 5 min.

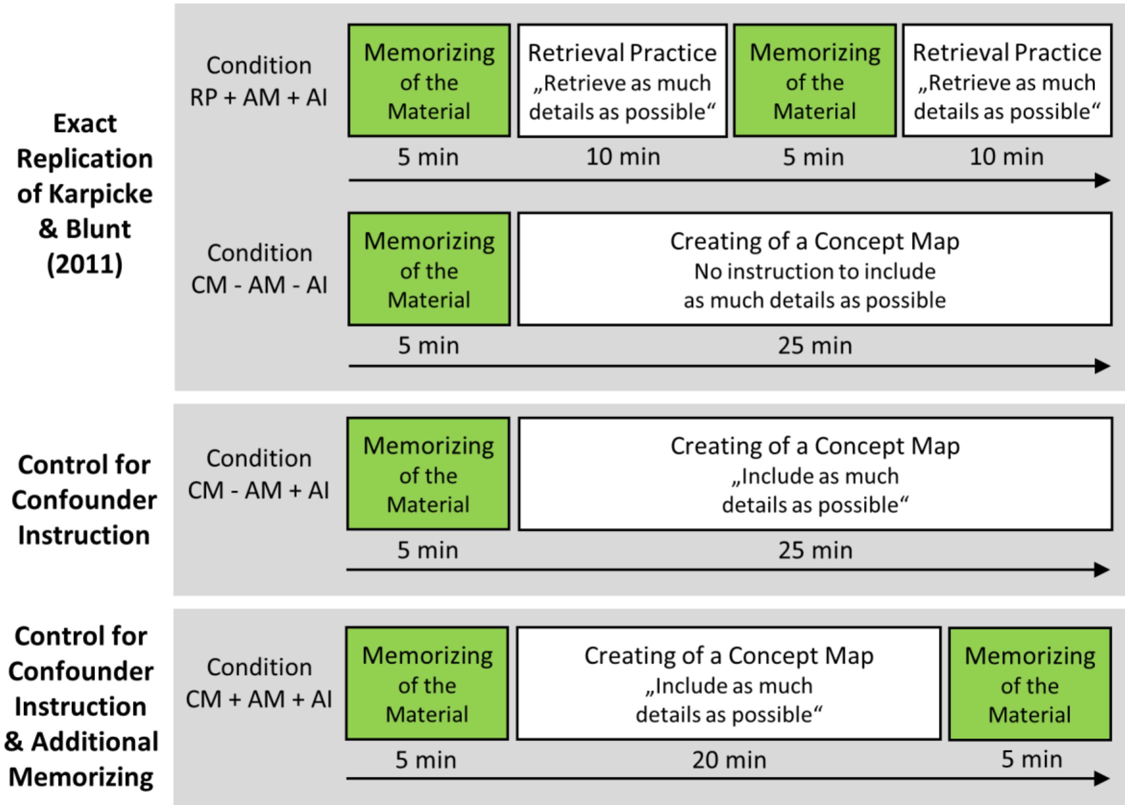

Figure 2. **Illustration of the four learning strategy conditions**. The "Retrieval Practice with Additional Memorization and with Additional Instruction 'Recall as Much as Possible' Condition" (RP + AM + AI) and the "Concept Mapping without Additional Memorization and without Additional Instruction 'Incorporate as Much as Possible' Condition" (CM - AM - AI) are exact replications of the conditions examined by Karpicke and Blunt (2011), i.e., RP + AM + AI = Karpicke and Blunt's "retrieval practice condition"; CM - AM - AI = Karpicke and Blunt's "concept mapping condition." In the "Concept Mapping without Additional Memorization and with Additional Instruction 'Incorporate as Much as Possible' Condition" (CM - AM + AI), to control for the confounder of different instructions, participants were prompted during the creation of the concept map as well that the concept map should contain as many details of the text as possible. In the "Concept Mapping with Additional Memorization and with Additional Instruction 'Incorporate as Much as Possible' Condition" (CM + AM + AI), to additionally control for the confounder of additional memorization, participants were asked to memorize the material after the creation of the concept map as well. Note that the text that the participants were to learn was available during the creation of the concept map but not during retrieval practice.

Afterward, in all four conditions, all participants filled out a questionnaire on metacognitive and demographic questions, which employed the very same items and scales as Karpicke and Blunt (2011).

The testing session, 1 week after the learning session, was identical for all four conditions: All participants were given the final test, i.e., the 14 verbatim and 2 inference questions. The time for the final test was not limited, which is identical to the procedure of Karpicke and Blunt (2011; Supporting Online Material).

Results

The proportion of correct answers for the verbatim questions and the inference questions in the final test as a function of experimental condition is shown in Figure 3 below.

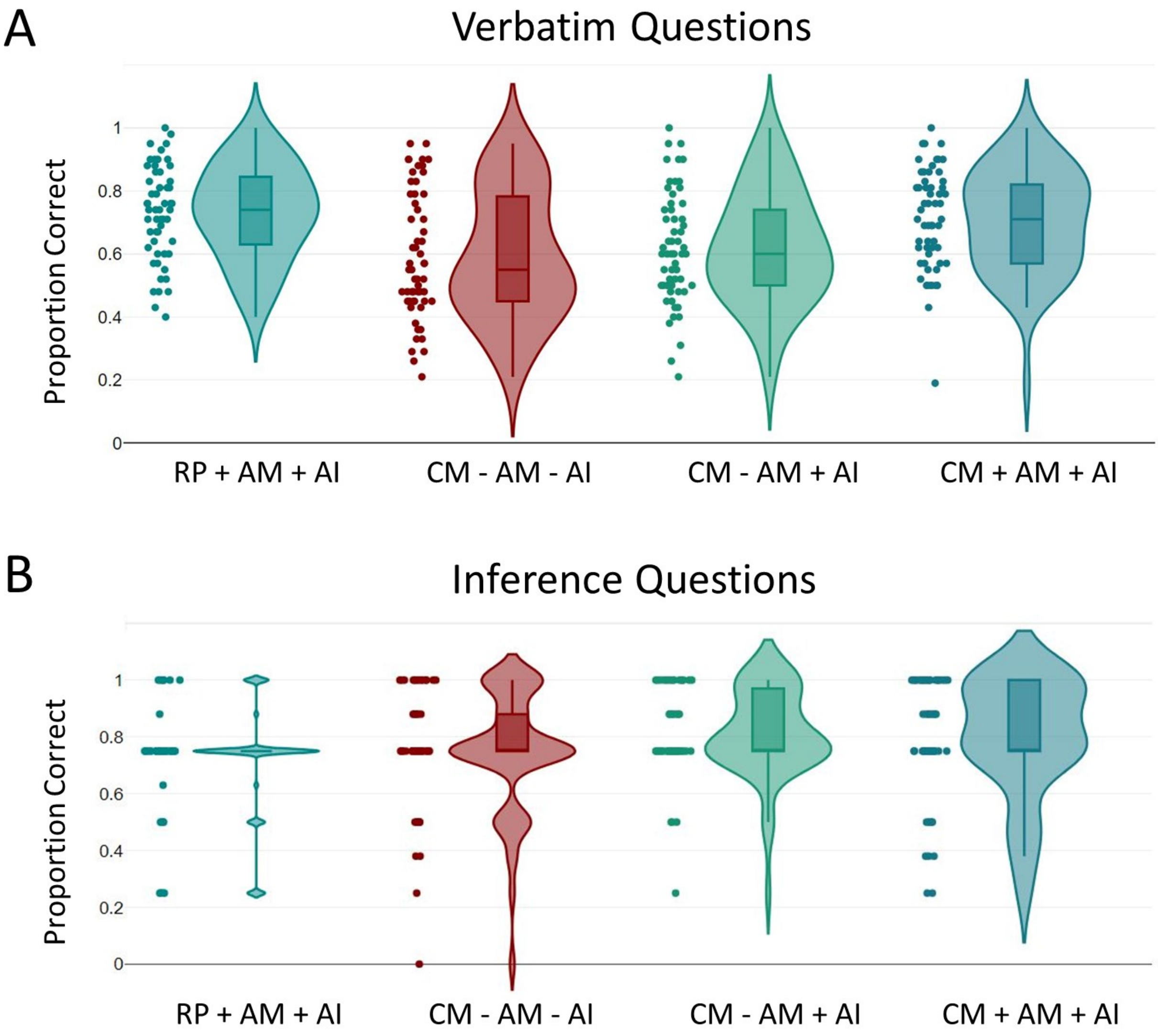

Figure 3. Memory Performance. The proportion of correct answers for verbatim questions (A) and inference questions (B) is shown as a function of the four learning strategy conditions (Retrieval Practice with Additional Memorization and with Additional Instruction "Recall as Much as Possible," RP + AM + AI; Concept Mapping without Additional Memorization and without Additional Instruction "Incorporate as Much as Possible," CM - AM - AI; Concept Mapping without Additional Memorization and with Additional Instruction "Incorporate as Much as Possible," CM - AM + AI; Concept Mapping with Additional Memorization and with

Additional Instruction "Incorporate as Much as Possible," CM + AM + AI). The violin plots show the probability density across participants; data points are plotted as dots. Center horizontal line markers show the medians. Box limits indicate the 25th and 75th percentiles. Whiskers extend 1.5 times the interquartile range from the 25th and 75th percentiles.

For the verbatim questions, an analysis of variance (ANOVA) with the factor of learning strategy condition (RP + AM + AI condition vs. CM - AM - AI condition vs. CM - AM + AI condition vs. CM + AM + AI condition) revealed a significant effect, $F(3, 226) = 8.33$, $p < 0.001$, $\eta_p^2 = 0.10$.

Post-hoc comparisons using the Tukey's HSD test indicated that performance was significantly higher ($p < 0.001$) in the RP + AM + AI condition ($M_{RP + AM + AI} = 0.73$, $SD = 0.15$) than in the CM - AM - AI condition ($M_{CM - ME - AI} = 0.59$, $SD = 0.20$). However, performance did not differ (p = 0.854) between the RP + AM + AI condition and the CM + AM + AI condition ($M_{CM + AM + AI} = 0.70$, $SD = 0.16$), indicating that the advantage of the retrieval practice condition disappeared when the same instruction was used in the concept mapping condition and when memorization time was equal. The CM - AM + AI condition ($M_{CM - ME + AI} = 0.62$, $SD = 0.18$) was outperformed by both the RP + AM + AI condition ($p = 0.004$) and the CM + AM + AI condition ($p = 0.041$). The CM – AM - AI condition was outperformed by the CM + AM + AI condition ($p = 0.003$) but not by the CM - AM + AI condition ($p = 0.842$), indicating that the instruction does not play a decisive role.[3]

For the inference questions, an ANOVA revealed a significant effect as well, $F(3, 226) = 2.99$, $p = .032$, $\eta_p^2 = .038$. Descriptively, performance was higher in all concept mapping conditions ($M_{CM - AM - AI} = 0.75$, $SD = 0.21$; $M_{CM - AM + AI} = 0.81$, $SD = 0.15$; $M_{CM + AM + AI} = 0.79$, $SD = 0.21$) compared to the retrieval practice condition ($M_{RP + AM + AI} = 0.71$, $SD = 0.19$). Post-hoc comparisons using the Tukey's HSD test indicated that the only statistically significant difference ($p = 0.038$) was between the RP + AM + AI and the CM - AM + AI condition. All other differences between conditions were not significant (all ps > 0.111). A comparison of the retrieval practice condition vs. all concept mapping conditions (collapsed data: $M_{CM} = 0.78$, $SD = 0.19$) showed that the performance in the retrieval practice condition was significantly lower, $t(228) = 2.41$, $p = 0.008$, $d = 0.36$.

[3] There was one extreme outlier in the CM + AM + RI condition (more than 3 SD below the mean; see also Fig.3). Removing this outlier, the results are as follows: $F(3, 225) = 9.17$, $p < .001$, $\eta_p^2 = .11$. The appropriate post hoc comparisons are: CM + AM + RI ($M_{CM + AM + RI} = 0.71$, $SD = 0.15$) vs. RP + AM + RI: $p = .950$; vs. CM - AM + RI: $p = .017$; vs. CM - AM - RI: $p < .001$).

To rule out that previous experience with the learning strategy, i.e., with retrieval practice or concept mapping, might have influenced the results, participants' previous experience was examined. The percentage of participants indicating that they had previous experience with the learning strategy they employed was higher in the retrieval practice condition (58.2%) compared to the concept mapping conditions (CM - AM - AI condition: 27.6%; CM – AM + AI condition: 34.5%; CM + AM + AI condition: 33.9%), $F(3,226) = 4.46$, $p = .005$, $\eta_p^2 = .056$. Previous experience with the learning strategy in the retrieval practice condition was higher compared to each of the individual concept mapping conditions (all *p*s < 0.043), which did not significantly differ from each other (all ps > 0.864), according to a Tukey's HSD post-hoc test. For both the verbatim and the interference questions, a four-by-two ANOVA with the between subjects factors of learning strategy (RP + AM + AI condition vs. CM - AM - AI condition vs. CM - AM + AI condition vs. CM + AM + AI condition) and previous experience with the learning strategy (previous experience vs. no previous experience) indicated neither a significant main effect of previous experience with the learning strategy (verbatim questions: $F(1, 222) = 0.12$, $p = .733$, $\eta_p^2 = .001$; inference questions: $F(1, 222) = 0.21$, $p = .651$, $\eta_p^2 = .001$) nor a significant interaction (verbatim questions: $F(3, 222) = 0.24$, $p = .871$, $\eta_p^2 = .003$; inference questions: $F(3, 222) = 1.74$, $p = .160$, $\eta_p^2 = .023$).

Furthermore, we examined previous knowledge about sea otters, assessment of text difficulty, and interest in the text to rule out that these factors may have influenced the results. There were neither statistically significant differences between the learning strategy conditions for previous knowledge on sea otters, $F(3, 226) = 2.44$, $p = .065$, $\eta_p^2 = .023$, nor for text difficulty, $F(3, 226) = 0.49$, $p = .690$, $\eta_p^2 = .006$, nor for interest in the text, $F(3, 226) = 0.03$, $p = .992$, $\eta_p^2 < .001$.

Concerning the judgments of learning, an ANOVA revealed a significant effect as well, $F(3, 226) = 10.22$, $p < .001$, $\eta_p^2 = .12$. Post hoc comparisons using the Tukey's HSD test indicated that judgment of learning in the RP + AM + AI condition ($M = 44.18$, $SD = 16.30$) was significantly lower than in the CM - AM - AI condition ($M = 54.10$, $SD = 16.86$; $p = 0.005$), the CM - AM + AI condition ($M = 59.14$, $SD = 17.09$; $p < 0.001$), and the CM + AM + AI condition ($M = 58.64$, $SD = 14.68$; $p < 0.001$). This replicates previous findings, showing that participants' assessment of how much they would remember 1 week later is significantly lower in the retrieval practice condition (e.g., Roediger & Karpicke, 2006; Karpicke & Blunt, 2011; but see Weissgerber & Rummer, 2023, for a critical discussion of judgements of learning in the context of retrieval practice).

Discussion

Our results clearly show that the memory advantage in the retrieval practice condition over the concept mapping condition reported in Karpicke and Blunt (2011) and, by extension, also in Lechuga et al. (2015), Camerer et al. (2018), and O'Day and Karpicke (2021), who employed the very same methodology, does in fact not prove that retrieval practice produces more learning than studying with concept mapping. When controlling for the methodological problem in these studies – namely that there was an additional memorization phase in the retrieval practice condition – the advantage of retrieval practice over concept mapping disappeared.

Concerning the verbatim questions, our data replicated Karpicke and Blunt's (2011) finding that performance in a retrieval practice condition where participants additionally memorize the learning material is better compared to a concept mapping condition without additional memorization. However, when participants also additionally memorize the learning material in the concept mapping condition, there is no statistically significant difference in performance between retrieval practice and concept mapping.

This finding indicates that Karpicke and Blunt's (2011) results were actually driven by the additional memorization in the retrieval practice condition rather than by differences inherent to the respective learning strategies, i.e., retrieval practice and concept mapping. The relevant role of memorization is further corroborated by the finding that performance in both conditions with additional memorization (RP + AM + AI and CM + AM + AI) was also better compared to the condition without additional memorization but where participants were instructed during the concept mapping task to cover as much information from the text as possible (CM - AM + AI). This represented another potential confounding factor in the study by Karpicke and Blunt (2011). The finding that performance in the concept mapping conditions without additional memorization did not differ as a function of the instruction provided during the concept mapping task indicates that the difference in the instruction does not play an important role for performance and is – at least in this setting – probably not a confounder.

Concerning the inference questions, the situation is entirely different from the verbatim questions. In contrast to Karpicke and Blunt (2011) – and to Lechuga et al. (2015) and O'Day and Karpicke (2021) as well – we unexpectedly found that the performance in the retrieval practice condition was lower than in the concept mapping conditions. As there were no significant differences in performance between the concept mapping conditions, neither the difference in the instruction nor – more importantly – in memorization seems to affect performance on the inference questions. However, from the perspective of classical test theory, measuring a highly complex construct such as meaningful learning with a diagnostic instrument consisting of merely two questions (or four scoring points) seems hardly adequate as very short test lengths negatively affect both reliability and validity (e.g., Novick, 1966; McDonald, 2013; Hogan, 2019). Thus, any conclusion drawn from such basis can only be tentative and must be taken with a pinch of salt.

In the present study, previous experience with concept mapping was lower than in Karpicke and Blunt's (2011) study. Lechuga et al. (2015) found that memory performance increased when participants were already familiar with and frequently used concept mapping compared to participants who had no experience in concept mapping and were trained for the purpose of the experiment. Accordingly, if the participants of the concept mapping condition in the present experiment had had a similar level of prior experience with concept mapping as in Karpicke and Blunt's (2011) study, their performance might have been even higher. In an applied context, this suggests that training in concept mapping and experience through regular application could improve performance, as already suggested by Mintzes et al. (2011).

The present study is mainly concerned with the methodology behind experiments comparing retrieval practice and concept mapping as learning strategies. However, the finding that the previously reported advantage of retrieval practice is actually driven by a confounder, i.e., by a different amount of memorization rather than by differences between the learning strategies of retrieval practice and concept mapping, has far-reaching consequences beyond methodology, which can only be touched upon here.

Concerning cognitive psychology, the advantage observed in previous studies of the retrieval practice condition over the concept mapping condition was explained by, for instance, the decisive role of better cue diagnosticity (Karpicke & Blunt, 2011) or active "access [to] already encoded information in memory" (Lechuga et al., 2015, p. 61). However, the present study now shows how the advantage of the retrieval practice condition observed in previous studies actually stemmed from additional memorization which was present in the retrieval practice condition but not in the concept mapping condition. Since the advantage of retrieval practice over concept mapping disappears when participants in the concept mapping condition, too, memorize, it seems to be the case that cognitive processes related to retrieval practice (such as cue diagnosticity or active access to already encoded information in memory) do not to improve memory, at least when studying textbook contents with elaborative learning strategies. In fact, this is in line with the results of a recent meta-analysis of the testing effect in classroom learning by Yang et al. (2021) who found virtually no advantage (Hedges' $g = 0.095$) of retrieval practice over various forms of elaborative learning strategies.

Concerning educational practice, the finding that the advantage of retrieval practice over concept mapping observed in previous studies is actually a methodological artifact challenges current recommendations for learning in real-life contexts. Based on their methodologically flawed findings, Karpicke and Blunt (2011), for instance, conclude that the human mind supposedly works in a way "that differs from everyday intuition" (p. 774) and that their finding may "pave the way for the design of new educational activities based on consideration of retrieval processes" (p. 774). In the light of the present findings, however, such conclusions seem invalid. When appropriately controlling for confounding factors in the previous studies, retrieval practice and concept mapping seem equally effective in promoting memory performance. However, it should be noted that the effectiveness of different learning strategies may vary as a function of the length of the retention interval, as suggested, for example, by the finding that the testing effect depends on the retention interval (e.g., Halamish & Bjork, 2011; Kornell et al., 2011; for a review, see Rowland, 2014). In Karpicke and Blunt's (2011) study as well as the present study retention intervals of 1 week were used so that equal effectiveness of retrieval practice and concept mapping, as observed in the present study, was demonstrated only for a retention interval of 1 week. Therefore, further research is needed to investigate whether the present findings also apply to other retention intervals.

The aim of the present study was to examine whether the memory advantage in the retrieval practice condition over the concept mapping condition, as observed in the paradigm developed by Karpicke and Blunt (2011), is actually not driven by retrieval practice itself but rather by the confounding variables of an additional memorization phase and the constantly visible instruction to retrieve as many details from the text as possible in the retrieval practice condition. The results clearly showed that the memory advantage observed in Karpicke and Blunt's (2011) paradigm indeed stems from these confounding variables because the advantage disappeared when the concept mapping condition also included – as was the case in the retrieval practice condition – an additional memorization phase and a constantly visible instruction to include as much information as possible from the text in the concept map. While the results of the present study clearly answered the research question for which it was designed, the results raise further questions for future research.

For instance, it is important to note that the additional memorization in the retrieval practice condition differed from the additional memorization in the concept mapping condition in one respect. In the retrieval practice condition, participants were asked to memorize the text again after retrieval practice, while in the concept mapping condition they were asked to memorize the concept map they had created. From an applied perspective, this makes sense because first studying the text by creating a concept map, but then putting that created concept map aside and then going back to the text to study for the upcoming test invalidates the idea of using the concept map to learn the text. Similarly, it would hardly make sense to provide participants in the retrieval practice condition with a concept map after retrieval practice and to ask them now to memorize the concept map instead of the text for the upcoming test. Therefore, from an applied perspective, it is important that the type of material memorized matches the appropriate learning strategy to ensure ecological validity.

However, from the perspective of basic experimental psychology, where the goal is to investigate basic cognitive mechanisms independent of applied contexts, it is interesting to see whether it makes a difference if participants additionally memorize either the text or the created concept map after having created a concept map. Interestingly, in a study by O'Day and Karpicke (2021), participants, after having created a concept map, performed a memorization task where they were asked to use the text for memorization and a retrieval task where they were asked to retrieve the contents of the text. The results of O'Day and Karpicke's (2021) Experiment 2, where the same concept mapping task was used as in our study, were fully consistent with the present results: Retrieval practice combined with additional memorization (so-called "retrieval practice" condition) only outperformed concept mapping when participants performed a concept mapping task without additional memorization and retrieval but not when participants additionally memorized and retrieved the text after the creation of the concept map. This learning activity, after having created the concept map, was a combination of text memorization and retrieval practice. Therefore, it is an interesting question for further basic research whether additional memorization of the text alone after a concept mapping task improves memory as well.

Similarly, it is important to note that the retrieval practice task and the concept mapping task differed in one aspect in the present study: in the retrieval practice task, the text the participants were to learn was not available, whereas, in the concept mapping task, the text was available. Again, from an applied perspective, this is reasonable because retrieval practice hardly makes sense when the text is available, or conversely, creating a concept map hardly makes sense when the text is not available. However, again from a basic experimental psychology perspective where research questions are not necessarily investigated with a focus on their applicability in real life, it would be interesting to examine what happens when retrieval practice is performed with the text being avaible, or conversely, when a concept map is created without the text being available. Indeed, the question of what happens when participants create a concept map without the text being available was already addressed in a previous study by Blunt and Karpicke (2014) and their results are fully consistent with the results of the present study. There, retrieval practice without the text being available in combination with additional memorization (so-called "retrieval practice" condition) did not outperform concept mapping without the text being available in combination with additional memorization (so-called "retrieval-based concept mapping" condition; Blunt & Karpicke, 2014).

These differences between the perspectives of applied and basic research, as presented in the preceding paragraphs, draw attention to the sometimes overlooked fact that the research logics of basic and applied research differ. Although the domains of real-life learning and experimental research overlap, their underlying rationalities diverge (e.g., Goldthorpe, 2001). From the perspective of basic experimental research, comparing specific learning conditions in isolation or comparing all possible combinations of learning conditions makes perfect sense, regardless of their relevance to applicability. However, such a research strategy does not necessarily make sense from the perspective of applied research as well because not all learning conditions that can be isolated or (re-)combined in different ways in the laboratory are feasible in real-life learning.

This case is illustrated in Figure 4 below. From a basic experimental perspective, the finding that (isolated) testing is more effective than (isolated) restudying is interesting and informative because it shows that different mental activities affect later memory performance differently. However, from an applied perspective, such a finding is less informative because in real-life learning, optimal studying actually comprises the combination of different learning strategies, including both testing and restudying, as reflected both in well-known study methods such as PQ4R (Thomas & Robinson, 1972) and in students' real-life learning behavior (Hartwig & Dunlosky, 2012; Blasiman et al., 2017; Kuhbandner & Emmerdinger, 2019). In particular, as illustrated in Figure 4 (on the right side), this problem may be obfuscated by the use of imprecise terminology. If the term "retrieval practice" is used to delineate a learning strategy which is actually a combination of retrieval practice and restudying, this may lead to results that may seem surprising and informative (e.g., "retrieval practice is better than restudying") at first glance, although they are actually rather trivial (e.g., "retrieval practice plus restudying is better than restudying alone"). Thus, potential implications for education drawn on the basis of experimental laboratory studies should be considered with caution as overemphasizing one factor or an oversimplified transfer to real-life learning may lead to already existing knowledge on learning being neglected.

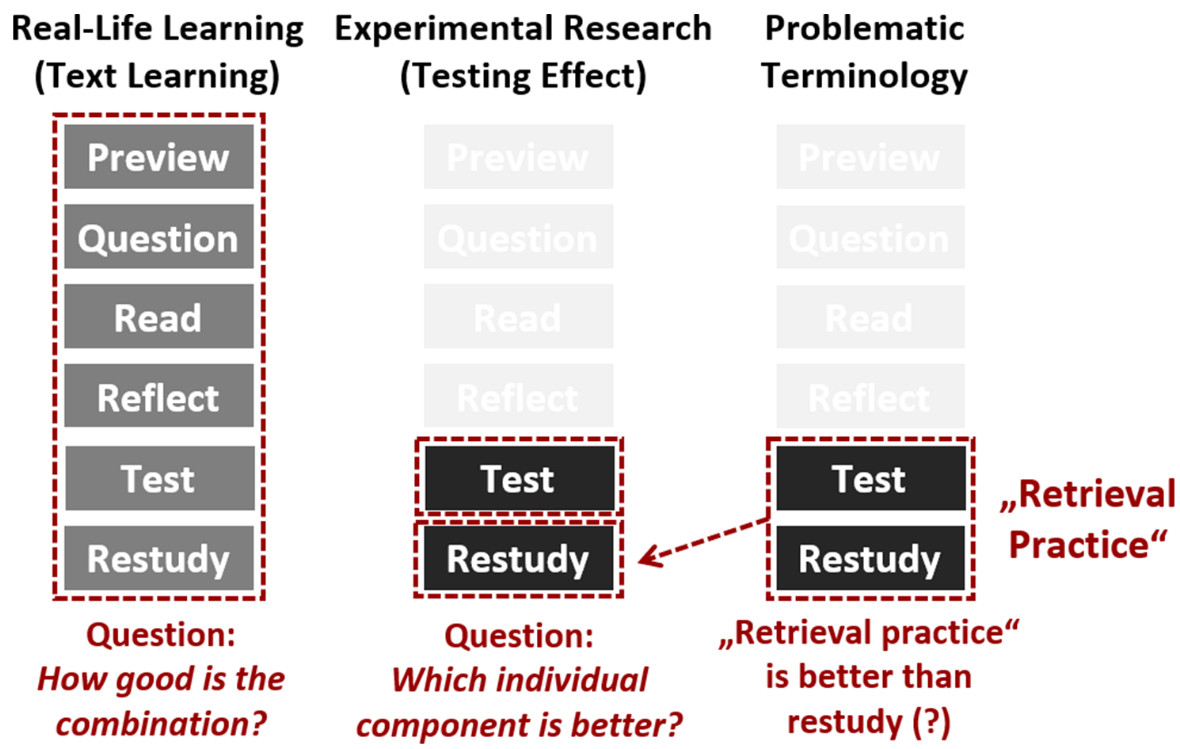

Figure 4. **Illustration of the divergent rationalities underlying real-life learning and experimental research.** Although the domains overlap, the focus of the questions asked is different: determining the optimal combination of cognitive processes (real-life learning) vs. determining the specific effect of isolated cognitive processes (experimental research). As shown on the right side, this problem may be obfuscated by the use of imprecise terminology. If the term "retrieval practice" is used to delineate a learning strategy which is actually a combination of retrieval practice and restudying, this may lead to results that may seem surprising and informative (e.g., "retrieval practice is better than restudying") at first glance, although they are actually rather trivial (e.g., "retrieval practice plus restudying is better than restudying alone"). Consequently, potential implications for education drawn on the basis of experimental laboratory studies should be considered with caution as overemphasizing one factor or an oversimplified transfer to real-life learning may lead to already existing knowledge on learning being neglected.

On a more general level, this study further demonstrates that it is essential in research to describe theoretical concepts and the related operationalizations in appropriate terminology. When investigating a complex topic such as learning strategies, which involve a variety of mental processes in different contexts, it is necessary to clearly define and delineate different learning strategies from one another so that unambiguous and valid conclusions can be drawn.

As shown in the present study, if the terms used to communicate a finding do not exactly reflect what participants actually did, invalid conclusions can be drawn. Although Karpicke and Blunt's (2011) retrieval practice condition included an additional second learning strategy, i.e., memorization, the authors did not account for this at the conceptual-linguistic level because they make general statements about retrieval practice and concept mapping as learning strategies. In other words, their terminology blurs and confuses what was actually done in their experiment. Thus, their conclusion that retrieval practice produces more learning than concept mapping – prominently featured in the title of their study – is both invalid and inaccurate in this generalized form and therefore misleading. In fact, similar problems at the level of terminology are found in other studies on retrieval practice as well, as shown, for instance in a recent study on the use of misleading terms in questionnaire studies on the use of retrieval practice in real-life learning (Kuhbandner & Emmerdinger, 2019).

In conclusion, by demonstrating that the advantage of retrieval practice over concept mapping observed in previous studies was actually driven by an additional memorization period in the retrieval practice condition, the present study serves as a reminder of the importance of a solid methodology. Furthermore, the present study also illustrates the importance of employing precise terms and language which precisely reflect – in both directions – the relation of theoretical concepts and actual operationalization. On a more general level, the present findings illustrate that one should be cautious when transferring experimental findings to real life learning contexts and be aware of the divergent rationalities underlying experimental research and educational practice.

3. General Discussion

3.1 Summary of the findings

Before the results of Studies 1-3 are brought together to form a synthesis that addresses the questions raised at the beginning, the results of the studies are first recapitulated individually.

In Study 1, psychology was examined as a scientific and academic discipline as a whole. Starting from the fundamental idea of what the subject of psychology actually is, it was first established that this question – in contrast to many other scientific disciplines – cannot be answered clearly and unambiguously. This is because the goal that psychology sets itself – namely to explain the human mind and behavior – covers the entire spectrum of human experience and behavior and is therefore not only extremely comprehensive but also highly disparate. For this reason, there is no thematic center of gravity in psychology that holds together the individual topics and areas studied in academic research, so that they cannot be concisely subsumed under a common unifying term. A comprehensive definition such as "human mind and behavior" is ultimately so broad that it covers (almost) everything and therefore (almost) nothing at the same time, and defining psychology as the "study of the psyche" is obviously circular. Psychology is thus difficult to distinguish thematically from other scientific disciplines from the social sciences and humanities, which also deal with people or specific aspects of human experience and behavior, even if one considers that psychology usually deals with the individual.

However, this lack of a unifying thematic bracket – in contrast to other academic disciplines that deal with similar topics – goes hand in hand with a strong emphasis on methodologies from the natural sciences, in particular the quantitative-experimental paradigm. The focus on a comparatively rigid methodology has far-reaching consequences, as this restriction means that psychological phenomena are primarily viewed and examined from a perspective that is shaped by the quantitative-experimental paradigm. This gives rise to two possible fundamental problems: First, aspects that cannot be grasped and measured within the quantitative-experimental paradigm may – intentionally or unintentionally – not be taken into account in psychological research. Second, the method is imposed on the aspects studied under the perspective of the quantitative-experimental paradigm, possibly changing their actual nature.

Study 1 ends with the consideration that the focus on a particular method represents a single-mindedness that is at odds with the broad thematic range of phenomena investigated by psychology, which is why greater methodological pluralism can help to improve the explanatory power of psychological research.

In Study 2, it was first demonstrated that the experiment as a research method is based on modern, particularly positivist, thinking and therefore also reflects the corresponding world view. On this basis, it was argued that despite this being rooted in modern thinking, the experiment necessarily also contains substantial elements that are not in agreement with modern thinking, but which instead can be better explained using postmodern approaches. Besides the theoretical considerations, such as research objectives and hypothesis generation, which form the foundation of any empirical study, it is primarily the observation that the experiment must necessarily be a reconstruction of naturally occurring phenomena if it is to investigate causal relationships, or, from a reverse perspective, that the experiment is neither something that occurs "naturally" nor an exact copy identical to the natural phenomenon. Instead, a large number of implicit and explicit assumptions flow into this reconstruction, which accordingly also influence the actual conduct of the experiment and its results. Therefore, in Study 2, it is argued that the experiment can be considered a simulacrum in the postmodern sense, especially if the results reflect an "artificial" experimental paradigm, that is, a situation in which the reconstruction of a naturally occurring phenomenon is very far removed from the original.

Going beyond the individual experiment, it was also investigated whether the results of experiments and their reception can also be interpreted according to postmodern approaches. The results of experiments can be absorbed by carrying out follow-up experiments in order to build on the original ones, extend them, transfer them to other contexts, or refute them. In addition, the results of this process can be communicated to the public beyond the scientific community if they are perceived as relevant in any way. Since experiments can be interpreted as simulacra, a series of experiments that relate to each other can, according to postmodern approaches, also be described as hyperreality, which refers to a system of interrelated simulacra. The main characteristic of a hyperreality is that such a system of connected simulacra not only appears to be as real as naturally occurring phenomena, but even more "real" than them, thus supplanting "actual reality". This interpretation of the experiment can be used to explain two prominent topics in psychology, namely the capacity of the visual working memory and the question of free will. Study 2 concludes with an outlook on the essential role of adequate terminology in psychological research and communication of its results in order to counter the problems outlined above.

In Study 3, a prominent finding was empirically investigated, namely the observation by Karpicke and Blunt (2011) that retrieval practice is a more effective learning strategy than concept mapping. This study was examined to determine whether the methodological, conceptual, and theoretical pitfalls found in Studies 1 and 2 influenced the design, the results, and their interpretation. It was found that Karpicke and Blunt's (2011) design contained two potential confounders that may have favored retrieval practice over concept mapping, namely, first, different memorizations times (two times five minutes in the recall condition versus one time five minutes in the concept mapping condition), and, second, different instructions for the participants. In order to test this assumption, an experiment was conducted in which these potential confounders were systematically varied. It turned out that the different instructions had no significant influence on the result but that the different memorization times actually functioned as a confounder that significantly influenced the result. For when learning took place two times for five minutes in both the recall condition and the concept mapping condition, the memory performance in the final memory test did not differ significantly between the two conditions and the advantage of retrieval practice over concept mapping disappeared.

Since performance in the memory test did not vary as a function of the learning strategy used but rather as a function of learning time, this result demonstrates that – contrary to Karpicke and Blunt's (2011) assumption – recall practice and concept mapping are equally efficient as learning strategies when the aim is to memorize facts from a simple text. From a cognitive psychology perspective, this result shows that the various explanations of cognitive mechanisms that are supposed to account for the purported advantage of retrieval practice over concept mapping are unfounded and cannot apply because this advantage is solely based on a methodological artifact. For educational psychology, this result means that retrieval practice is not a better learning strategy than concept mapping, so recommendations to implement more retrieval practice in learning are unfounded. Finally, Study 3 argues that Karpicke and Blunt's (2011) study exemplifies how mixing the different research logics of basic and applied research can lead to problematic results.

3.2 The experiment at the intersection of theory, methodology, and language

Psychology as a science – like all other sciences – has developed historically and is therefore also a product of academic institutionalization (e.g., Greenwood, 2015; Hothersall, 2022; King et al., 2013; Leahey, 2018; Valsiner, 2012). Therefore, it is important to recognize that the way psychological research is conducted at any given time is neither the only possible way to conduct research, arrive at insights, and generate knowledge, nor necessarily the best way from an epistemological perspective, all because it is legitimized by the weight of tradition.

For reasons beyond the scope of this study, the experiment has assumed a prominent role in psychological research. This alone would be cause enough to analyze the experiment from a philosophy of science perspective. However, there is an even more fundamental reason for such an analysis, which lies in the object of investigation and the goals of psychology because the methods used in a scientific discipline must also be suitable for adequately capturing, describing, and explaining the object of investigation. But what does "adequate" actually mean in this context? The object of investigation and the goals of a discipline are not determined a priori and from the outside – decreed (by whom?) in an official act of foundation, as it were – but rather develop inextricably linked together with the discipline and out of it. Thus, as Feyerabend (1975; see also Farrell, 2003; Niaz, 2020; Oberheim, 2006; Preston, 1997) in particular has pointed out, there is no objective criterion by which one can measure whether the phenomena studied by psychology are adequately captured and explained. This is because in order to be able to clearly and objectively assess such adequacy it would be necessary to have reliable and definite knowledge about the phenomenon in question in advance – thus anticipating precisely the desired result of the research, which in turn would render research itself superfluous.

However, since at the same time no discipline – especially not in institutionalized form – can function without reviewing and justifying the knowledge it generates, there are two basic ways of doing this, both of which can be used alongside each other: First, an approximate assessment of adequacy can be made retrospectively, but this faces the same problem of which criteria to use and how to justify them. Second, assessment criteria can be established in advance, and if these are met, adequacy is practically automatically assumed, so to speak. The second option basically takes up the goal of logical empiricism (e.g., Achinstein & Barker, 1969; Richardson & Uebel, 2007; Uebel & Limbeck-Lilienau, 2022) or critical rationalism (e.g., Keuth, 2005; Rowbottom, 2011) to create binding guidelines for what is considered valid science and research. In particular, this second option is associated with the institutionalization of research, as such a fixed program provides a supportive framework within which research can be organized according to clear rules. On the one hand, this has the advantage that such institutionalization and professionalization are usually associated with access to a wide range of resources and security for researchers, but on the other hand it is usually accompanied by an increasing marginalization of dissenters and their ideas who do not fit into this pattern, as Kuhn (1962) and Lakatos (1978; see also Jagtenberg, 1983; and especially for psychology, see Lunt, 1999) in particular have pointed out.

This last point is particularly relevant for the present study, because – as shown in Study 1 – psychology is largely characterized by the quantitative-experimental method while other approaches not only play a more marginal role but are also often considered less substantial. Thus, the experiment represents the framework within which research in psychology is often conducted, which – in light of the fact that other approaches are marginalized – leads to far-reaching consequences for the object of investigation of psychological research. Because if the experiment is the method of choice that is resorted to first, as the default, as it were, then the research process does not take its starting point from the phenomenon under investigation itself, to which access is sought in order to understand it. Instead, the method – i.e., the experiment – constitutes the starting point, and this approach establishes from the outset a certain framework and boundaries within which the phenomenon in question is approached and within which the phenomenon is ultimately *re*-constructed, as discussed in Study 2. With such an approach, the phenomenon is almost inevitably no longer perceived "as it is" and naturally occurs but rather from the perspective of the experiment, emphasizing certain aspects that may be more accessible to the experiment than others, or transferring the entire phenomenon into something that can be studied experimentally.

However, since the object of investigation in psychology is the entire spectrum of the human mind and behavior – i.e., an enormous, almost unlimited number of phenomena – the (self-)limitation to the experiment as a method restricts and reduces this richness to what is experimentally accessible. This implies two possible underlying implicit presuppositions and raises two questions: Such an approach therefore presupposes that either the (almost) entire spectrum of the human mind and behavior can be investigated experimentally or that only those phenomena are relevant that can be investigated experimentally. Although both presuppositions go beyond the scope of the present study and require their own comprehensive discussion from the perspective of scientific theory, especially epistemology, it should be noted here that neither option can be taken for granted, particularly because they are very strong assumptions.

The first question that arises here concerns the object of research in psychology because the restrictions due to the focus on the experiment mean that it is unclear whether it can actually be said that psychology studies the human mind and human behavior generally and comprehensively or whether it does not rather (almost) only study certain – namely experimentally accessible – aspects of it. Depending on the answer, the self-conception of psychology as an academic discipline would be more or less affected.

The second question concerns the way in which psychology understands and implements science. For if the goal of science is to discover universal laws (e.g., Bonnay, 2018), and if psychology shares this goal, it should strive to discover such universal laws that allow the enormous variety of phenomena of human mind and behavior to be reduced to and explained by a smaller number of more fundamental mechanisms. Psychology has certainly generated a large number of more or less fundamental theories. However, given the observation that many findings in psychology reflect an approach to phenomena that is limited by the focus on the experiment as a method, the question arises as to whether these fundamental theories actually describe fundamental mechanisms. For it is equally possible that this reduction of a multitude of phenomena to a smaller number of fundamental mechanisms is rather based on the fact that from the entire range of all possible phenomena, only those that can be experimentally investigated were selected and examined. In other words, the reduction of the number of phenomena that can be explained by psychological theories is not necessarily due to the fact that these theories have so much explanatory power that they can cover many phenomena and explain them on the basis of a few fundamental mechanisms. Instead, however, it is also possible that only those phenomena were selected and investigated that were accessible to the experimental method and that these results were then unduly (over-)generalized. In this case, the explanatory power of psychological theories would be overestimated, and in part illusory, because phenomena that are accessible by experiment are erroneously considered representative of a larger number of phenomena that may not be accessible by experiment.

From a postmodern perspective, as adopted in Study 2, this can also be expressed in terms of simulacrum and hyperreality: The limitations resulting from the choice of the experiment as a method of investigation mean that the *re*-construction of naturally occurring phenomena in the experiment can be seen as a simulacrum. If universal laws are derived from such experiments – i.e., simulacra – as is the aim of science, the laws discovered in this way can be interpreted as hyperreality for two interrelated reasons: First, general laws are usually not derived from a single experiment, but from a series of interconnected experiments, and such a system of experiments – as simulacra – can be regarded as hyperreality, as explained in Study 2. Second, and this reason is crucial here, hyperrealities are characterized by the fact that they are perceived as "more real" than the actual reality, although they usually deviate from it because they are permeated by distortions. With respect to deriving universal laws from the experiment, this means that the laws found in this way are not necessarily actually (universally) valid, but are instead based on a limited selection of cases – cases that can be investigated in the experiment. However, the results obtained in this way are generalized and regarded as universally applicable and valid, although they only describe specific, narrowly defined situations. The issue here is therefore not whether the laws derived in this way are actually universally valid, as this is a question that must be examined in each individual case, or the fundamental philosophical problem of their ontic status. Instead, it is a matter of demonstrating that such universal laws established with the help of experiments, as a result of the way in which they come about, harbor the danger of not being universally valid but merely reflecting the specific situations under which they came about.

This interpretation as simulacra and hyperreality is illustrated below, using the subjects examined in Studies 2 and 3: For the visual working memory, the very general and far-reaching rule has been established, based on experimental research, that its capacity is approximately four items (Luck & Vogel, 1997). This finding quickly became widespread and was passed on in textbooks (e.g., Baddeley, 2007; Parkin, 2013; Goldstein, 2015) and researched in further experiments (see the summaries by Brady et al., 2011; Luck & Vogel, 2013; Ma et al., 2014; Schurgin, 2018). What matters in this study is that the finding of four items as the capacity of the visual working memory was widely accepted as a universal law that could adequately describe a very wide range of phenomena – namely the entire visual perception, i.e. a very rich world – although the original experiments on which this law was based only investigated a very specific situation, namely simple colored geometric shapes against a gray background. This situation has the characteristics of a hyperreality because a "universal" law that reflects and generalizes the highly restricted *re*-construction of a naturally occurring phenomenon was considered a reality, although the capacity of the visual working memory is much greater in actual real-life situations.

Similarly, the experiment by Libet et al. (1983) was seen as proof that there is no free will and that all our decisions are already determined by neuronal mechanisms before we even become aware of them (e.g., Gazzaniga, 2011; Wegner, 2017; see Kihlstrom, 2017, for further examples). However, this extremely far-reaching universal law was solely derived from an extremely limited experimental situation, namely the moment when a finger is moved. This interpretation and subsequent generalization of a very specific situation was often accepted as "real" (but also met with resistance; for a summary of this complex discussion, see Nahmias, 2010; Radder & Meynen, 2013; Schlosser, 2014; Fischborn, 2016; Lavazza, 2016; Schurger, 2017). The hyperreal character of this process is evident in the fact that, on the basis of an experiment in which an extremely broad concept such as free will, which encompasses a wide range of different phenomena, was investigated, precisely this perception, which is accessible in everyday experience and according to which free will manifests itself in a wide variety of situations, is negated and instead an extremely limited and highly artificial situation – the movement of a finger under laboratory conditions – is considered to be more relevant and meaningful, i.e., more "real". This observation does not imply a statement as to whether free will exists or not but rather that the direct personal experience of being able to act freely is masked and overridden by such a simulacrum. At this point, a further aspect becomes relevant, which again suggests how much the implementation and interpretation of experiments depends on presuppositions, because people with a materialistic world view are presumably more willing to accept the results and above all the interpretation of Libet et al. (1983) as a refutation of free will, because this confirms their attitude – also in the sense of confirmation bias (e.g., Nickerson, 1998; Wason, 1960 & 1968).

The experiment by Karpicke and Blunt (2011) examined in Study 3 can also be interpreted as hyperreality, because there – similar to Libet et al. (1983) – a comprehensive and complex topic, namely learning strategies and meaningful learning, was implemented in a reduced and artificial way that bears little resemblance to how students actually learn (Hartwig & Dunlosky, 2012; Blasiman et al., 2017; Kuhbandner & Emmerdinger, 2019). However, the result, which was also distorted by a confounder, was taken up in further studies with the same design (Camerer et al., 2018; Lechuga et al., 2015; O'Day & Karpicke, 2021) and researched further, resulting in a system of interconnected simulacra. In addition, the results obtained in this way, which reflect an artificial situation, were recommended for implementation in actual learning situations outside the laboratory, i.e., artificial phenomena were superimposed on real situations – or in other words: a hyperreality was created.

The problem of deriving universal laws on the basis of limited and artificial phenomena that are reconstructed in experiments is also evident at the linguistic-conceptual level and is often exacerbated by the language used. Thus, the experiment by Luck and Vogel (1997) refers to the capacity of the visual working memory in general, although in fact only the capacity of the visual working memory for simple geometric shapes against a gray background was examined. Libet et al. (1983) spoke about free will, but this actually only meant moving a finger within a narrow time window. In Karpicke and Blunt's (2011) experiment, it was stated that retrieval practice and concept mapping were to be compared as learning strategies in terms of meaningful learning, but only one component of this, namely recall, was measured under artificial conditions.

These three studies have in common that only one very specific phenomenon was investigated, but the results reflecting this particular situation were then generalized linguistically into a universal law. The "universal law" obtained in this way is therefore not based on an abstraction in which many different phenomena have been traced back to a common mechanism after extensive empirical testing, and which can accordingly also explain a large number of different phenomena. Instead, abstraction only takes place on a linguistic level, where a specific phenomenon is considered representative of a whole class of phenomena and a general umbrella term is used accordingly, although this umbrella term would have to be restricted according to what was actually investigated in the experiment. Phenomena of different categories and levels of abstraction are thus linguistically equated without this equating necessarily being empirically justified.

Language plays a decisive role not only in establishing universal laws, but also much earlier in the research process, namely during the conception and planning of the experiment. This is because the question that the experiment seeks to answer is based on a corresponding theory that makes assumptions about causal relationships in the phenomenon under investigation (see, e.g., Craver & Tabery, 2023; Glennan, 1996 & 2002; Kauffman, 1970). The theory therefore makes it possible to identify potentially causal factors that may affect the phenomenon in question. This identification, coupled with considerations about how which changes in the potential causal factors could affect the phenomenon, thus forms the necessary theoretical foundation on which every experiment is based, because without this theoretical foundation, any experimental investigation would be purely accidental. The result obtained by the systematic variation of potential causal factors then enables the empirical verification of these predictions and thus provides the criterion for deciding whether the previously made assumptions are correct or not. This procedure is thus decisively influenced by the theory used and thus also by language, because a theory represents a system of interconnected concepts, and both the concepts and their interaction are expressed in language. Theory and language thus inextricably belong together and form the foundation of any experiment, and this point also reinforces the reasoning of Study 2, according to which experiments are fundamentally and inevitably a postmodern endeavor.

The theoretical groundwork that forms the basis of the experiment is itself influenced by various factors, including specific assumptions or prior knowledge about the phenomenon in question or more general worldviews, as shown in Study 2. In addition, however, the understanding of science also plays an important role, and if, as outlined in Study 1, the experiment is regarded as the most important or even (almost) exclusive approach to phenomena, there is obviously an inherent tendency to approach each phenomenon from the perspective of the experiment, as it were by default, which means that the theoretical foundation will also have the task of enabling and shaping this approach in a meaningful way. This confirms the reasoning of Study 2, according to which an experiment is the reconstruction of a naturally occurring phenomenon, a reconstruction that – evidently – happens from the perspective of the experiment.

Even from this brief sketch it is clear that theory and language permeate the entire experiment and are therefore of decisive importance as to whether or not an experimental investigation leads to an increase in knowledge – or, on the contrary, rather creates confusion or unduly invalidates existing knowledge. Phenomena, questions, theories, and concepts are of course closely related but they are obviously not identical, which is why there is a risk that they will be confused and conflated, and this can also distort the experiment and its results. Therefore, as outlined in Study 2 and 3, it is essential that the language and terminology used in research is, first, clear and unambiguous, and, second, adequately reflects the phenomenon in question, the research question, the underlying theories, and the concepts used. The level of abstraction is particularly relevant here, as shown in Study 2 for the capacity of the visual working memory and the Libet-type experiments on free will, where a very specific and limited experimental situation was overgeneralized to very far-reaching universal laws.

The critical role of language and the concepts it reflects can also be demonstrated for the comparison of recall practice and concept mapping as learning strategies with regard to meaningful learning by Karpicke and Blunt (2011), as examined in Study 3, because there an unclear terminology led to a questionable operationalization so that the results must be interpreted as a methodological artefact. This point can be illustrated by the unclear and fuzzy use of the term "study(ing)", which results in far-reaching problems. In Kuhbandner und Emmerdinger (2019), it was shown that crucial terms regarding learning, studying, and learning strategies were often used inconsistently, ambiguously, or indiscriminately – and therefore unclearly – in previous studies. The same ambiguity also affects the methodology of Karpicke and Blunt's (2011) experiment, as shown by the following quotes, where the same activity is indiscriminately referred to by "study" or "(re-)read": "Students in all conditions *studied* the text in an initial 5-min *study* period" (supporting online material, p. 1; all italics by RM); "[i]n the elaborative concept mapping condition, after *reading* the text in the initial 5-min *study* period" (p. 1); "[i]n the retrieval practice condition, after *studying* the text for 5 min" (p. 2); "[s]ubjects then *reread* the text in another 5-min *study* period" (p. 2). Apart from the fact that (re-)reading is only one way of studying learning material, i.e., it is a subcategory of studying here, "study" is at the same time also used in Karpicke and Blunt (2011) in an overarching sense, since a learning strategy as a whole is likewise designated by this term, namely, "elaborative *studying* with concept mapping". By contrast, retrieval practice is not referred to as "studying", although it is compared to concept mapping at the same level as a learning strategy. Similarly, everything that happens in the experiment in the context of the particular learning strategy is indiscriminately referred to as "learning" or "studying", as the following quotation shows: "The students first *studied* a science text under one of four conditions within a single initial *learning* session" (Karpicke & Blunt, 2011, p. 772). These examples clearly demonstrate the need for a precise terminology.

The bottom line of this chapter is thus the observation that – as the three examples of the capacity of the visual working memory, free will, and the comparison of learning strategies with regard to meaningful learning have shown – the use of the experiment as a method of gaining knowledge may be accompanied by major problems and can therefore actually hinder the goal of generating reliable and relevant knowledge. Contrary to the widespread attitude in academic psychology, according to which the experiment is by default, as it were, regarded as the best and often the only way to gain reliable knowledge, this study argues that if the fundamental epistemological characteristics of the experiment are not sufficiently taken into account, results are obtained that do not systematically allow adequate conclusions to be drawn about the phenomenon in question. Instead, there is even a risk that this will generate knowledge about a phenomenon that not only does not expand the understanding of this phenomenon, but also negatively affects the existing understanding if inaccurate assumptions are regarded as true.

In a nutshell: The fit between the phenomenon in question as the object of investigation and the experiment as the method of investigation used to approach the phenomenon is therefore a factor that plays a major role in determining the results of an academic discipline.

3.3. The future of psychology

The present study is based on the observation that psychology as an academic discipline is, on the one hand, thematically an extremely broad discipline that encompasses a multitude of phenomena so that its object of investigation is very vaguely described as the totality of the human mind and behavior, which inevitably results in numerous overlaps with other scientific disciplines. On the other hand, academic psychology – in contrast to other disciplines such as history, which draws on a wide variety of methods to approach its similarly broad object of investigation – is characterized by a comparatively narrow spectrum of methods, namely a dominance of the experiment as the method of choice for generating knowledge. This relationship was described by Jüttemann (1983) as "principle of inversion".

This importance of the experiment has far-reaching consequences for the knowledge generated in psychology because the emphasis on the experiment leads to other ways of approaching phenomena and generating knowledge being marginalized. This is problematic for two reasons: First, this limits the scope of phenomena investigated by psychology, because either only those phenomena are considered that are suitable for experimental investigation, those aspects of phenomena that are experimentally accessible are emphasized, or the phenomena are reconstructed in an experimental investigation in such a way that they only reflect the naturally occurring phenomenon – the actual object of investigation – in a distorted way. In other words: Naturally occurring phenomena are forced onto the proverbial Procrustean bed, with some fitting neatly into the bed, others too large for it and therefore curtailed, and still others too small and thus forcibly stretched to fill the frame. Second, the examples examined in this study show that the epistemological and methodological pitfalls inherent in the experiment are by no means so obvious as to be easily avoided. In both cases it is doubtful whether the knowledge generated in this way is reliable, in the sense that it tells us something about naturally occurring phenomena, or whether a hyperreality is manufactured that creates the impression that reliable knowledge is being generated, but which actually only revolves around itself.

The pitfalls of the experiment as a method include three interrelated aspects in particular: First, there is a danger that phenomena, research questions, theories and concepts will be conflated or remain unclear. Second, this can mean that the theoretical foundation on which the concrete empirical investigation in the form of the experiment is then based is inadequate. Third, language – and thus conceptual thinking – necessarily permeates all aspects of the experiment, from the theoretical foundation to the implementation and interpretation of the results. There is the danger, as shown in Studies 2 and 3, that very limited and specific situations – which in turn often reflect the inherent limitations of the experiment – are overgeneralized via language without the empirical findings warranting this.

On the one hand, these potential problems can compromise the transferability of an artificial situation in the laboratory, in which naturally occurring phenomena are reconstructed in a reduced form, to real situations with all their richness, diversity, and also ambiguity. On the other hand, it is therefore also debatable whether it is justified to derive universal laws from such limited and specific situations. The extent to which the potential problems identified in the present study affect the actual output of psychology cannot be assessed here as this would require extensive research, but the three prominent examples of visual working memory capacity, free will, and learning strategies examined in this study suggest that this danger is by no means negligible.

The reasons why these problems associated with the experiment as a method presumably permeate larger parts of academic psychology lie outside the scope of this study, but a possible explanation is outlined here: As explained above, psychology, like any other science, is a social endeavor, which is why social factors such as pressure to conform (for social pressure in science, see Mulkey, 1970 & 1976; Shibayama & Baba, 2015) or being part of hierarchies (for hierarchies in science, see Chafetz & Fox, 2006; Elias et al., 1982), especially in conjunction with cognitive factors such as availability heuristics (e.g., Reber, 2017; Tversky & Kahneman, 1973 & 1974), may play a role here. Therefore, the use of the experiment as a method and its prominent position may be taken for granted without question, without dealing with the associated problems or looking for alternatives.

So what does it mean for psychology as an academic discipline if, for all these reasons, the fit between the naturally occurring phenomenon we are interested in and the way in which this phenomenon is studied and knowledge is generated about it is lost? One possibility would be to discuss the question of whether the experiment is actually the best method for gaining reliable knowledge about the subject of psychology, namely the human mind and behavior. This also implies the fundamental question that characterizes psychology as a scientific discipline as to whether the reduction of the enormous range of phenomena of the human mind and behavior to a few explanatory patterns is actually desirable or whether it is a matter of ignoring and neglecting the richness of the world and human experience, which might limit the potential of the human being.

In this context, two recent developments should be mentioned that deal with similar questions as the present study. First, various suggestions have been made as to how a stronger pluralism of methods could be implemented and how psychology could benefit from this (e.g., Hutmacher & Mayrhofer, 2023; Malich & Rehmann-Sutter, 2022; Wiggins & Christopherson, 2019). Second, as has been suggested in the context of the replication crisis in particular, would be to improve the theories on which empirical, especially experimental, research is based (Fiedler, 2017 & 2018; Scheel et al., 2021; Scheel, 2022). On the one hand, this implies that the theories should more adequately reflect the phenomena they address and postulate causal mechanisms with actual explanatory power. On the other hand, this also involves a shift in focus away from empirical research, so that theory provides the foundation and is therefore on an equal footing with empirical research. It is therefore possible, as outlined by Morawski (2019), to resort more to philosophy of science to provide answers to the fundamental questions raised by the replication crisis.

As the replication crisis suggests, psychology as a scientific discipline is at a crossroads and the way in which the questions raised in this study are answered is likely to affect the direction in which psychology will develop. Will psychology mainly revolve around itself and focus on the phenomena it itself creates? Then there is a danger that psychology will lose its relevance because the knowledge it generates loses touch with what actually moves humanity. Or will psychology instead shed light on the phenomena that actually make up the human inner world? Psychology can then make an important contribution to shaping the future of our society through a better understanding of humankind.

References

Achinstein, P., & Barker, S. F. (Eds.). (1969). *The legacy of logical positivism*. The Johns Hopkins Press.

Agassi, J. (1977). *Towards a rational philosophical anthropology*. Nijhoff. https://doi.org/10.1007/978-94-010-1095-5

Agassi, J. (1981). *Science and society. Studies in the sociology of science*. Reidel.

Agra, G., Formiga, N. S., Oliveira, P. S. D., Costa, M. M. L., Fernandes, M. D. G. M., & Nóbrega, M. M. L. D. (2019). Analysis of the concept of meaningful learning in light of the Ausubel's theory. *Revista brasileira de enfermagem, 72*, 248–255.

Andersen, H., & Hepburn, B. (2015). Scientific method. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/scientific-method/

Armstrong, D. M. (1997). *A world of states of affairs*. Cambridge University Press. https://doi.org/10.1017/CBO9780511583308

Ausubel, D. P. (1963). *The psychology of meaningful verbal learning. An introduction to school learning*. Grune & Stratton.

Ausubel, D. P. (1968). Facilitating meaningful verbal learning in the classroom. *The Arithmetic Teacher, 15*(2), 126–132. https://doi.org/10.5951/AT.15.2.0126

Aylesworth, G. (2015). Postmodernism. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy*. https://plato.stanford.edu/entries/postmodernism/

Baddeley, A. (2007). *Working memory, thought, and action*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780198528012.001.0001

Barthes, R. (1963). L'activité structuraliste. In R. Barthes, *Essais Critiques* (pp. 215–218). Éditions du Seuil. [Structuralist activity. In R. Barthes, *Critical Essays* (pp. 213–220), translated by R. Howard, (1972). Northern University Press].

Baudrillard, J. (1981). *Simulacres et simulation*. Galilée. [*Simulacra and simulation*. Translated by S. F. Glaser, (1994). The University of Michigan Press. https://doi.org/10.3998/mpub.9904]

Baumeister, R. F. (2019). Self-control, ego depletion, and social psychology's replication crisis. In A. Mele (Ed.), *Surrounding Self-control* (pp. 28–46). Oxford University Press. https://doi.org/10.31234/osf.io/uf3cn

Bazan, A. (2018). Psychoanalysis and academia: psychoanalysis at the crossroads between exact and human sciences. *International Forum of Psychoanalysis, 27*, 90–97. https://doi.org/10.1080/0803706X.2017.1392040

Beebee, H. (2013). *Free will: An Introduction*. Palgrave Macmillan.

Bem, S., & De Jong, H. L. (2013). *Theoretical issues in psychology. An introduction*. Sage. https://doi.org/10.4135/9781526402097

Bergmann, G., & Spence, K. W. (1941). Operationism and theory in psychology. *Psychological Review, 48*(1), 1–14. https://doi.org/10.1037/h0054874

Bermúdez, J. L. (Ed.). (2006). *Philosophy of psychology. Contemporary readings*. Routledge. https://doi.org/10.4324/9780203028957

Bertens, H. (1995). *The idea of the postmodern. A history*. Routledge. https://doi.org/10.4324/9780203359327

Bird, A. (2013). The historical turn in the philosophy of science. In M. Curd & S. Pillos (Eds.). *The Routledge companion to philosophy of science* (pp. 79–89). Routledge. https://doi.org/10.4324/9780203744857

Blasiman, R. N., Dunlosky, J., & Rawson, K. A. (2017). The what, how much, and when of study strategies: Comparing intended versus actual study behaviour. *Memory, 25*(6), 784–792. https://doi.org/10.1080/09658211.2016.1221974

Blunt, J. R., & Karpicke, J. D. (2014). Learning with retrieval-based concept mapping. *Journal of Educational Psychology, 106*(3), 849–858. https://doi.org/10.1037/a0035934

Bonnay, D. (2018). Scientific explanation. In A. Barberousse, D. Bonnay, & M. Cozic, (Eds.), *The philosophy of science. A companion* (pp. 3–52). Oxford University Press. https://doi.org/10.1093/oso/9780190690649.003.0001

Bourdieu, P. (1984). *Homo academicus*. Les Éditions de Minuit. [*Homo academicus*. Translated by P. Collier, (1990). Polity Press.]

Botterill, G., & Carruthers, P. (1999). *The philosophy of psychology*. Cambridge University Press. https://doi.org/10.1017/CBO9780511612428

Brady, T. F., Konkle, T., & Alvarez, G. A. (2011). A review of visual memory capacity: beyond individual items and toward structured representations. *Journal of Vision, 11*(5), 1–34. https://doi.org/10.1167/11.5.4

Brady, T. F., Störmer, V. S., & Alvarez, G. A. (2016). Working memory is not fixed-capacity: more active storage capacity for real-world objects than for simple stimuli. *Proceedings of the National Academy of Sciences, 113*(27), 7459–7464. https://doi.org/10.1073/pnas.1520027113

Brass, M., Furstenberg, A., & Mele, A. R. (2019). Why neuroscience does not disprove free will. *Neuroscience & Biobehavioral Reviews, 102*, 251–263. https://doi.org/10.1016/j.neubiorev.2019.04.024

Bucchi, M. (2002). *Scienza e società. Introduzione alla sociologia della scienza*. Mulino. [A. Belton (Transl.). (2004). *Science in society. An introduction to social studies of science*. Routledge. https://doi.org/10.4324/9780203299739]

Bugelski, B. R. (1962). Presentation time, total time, and mediation in pairedassociate learning. *Journal of Experimental Psychology, 63*(4), 409–412. https://doi.org/10.1037/h0045665

Bunge, M., & Ardila, R. (1987). *Philosophy of psychology*. Springer. https://doi.org/10.1007/978-1-4612-4696-1

Camerer, C. F., Dreber, A., Holzmeister, F., Ho, T. H., Huber, J., Johannesson, M., Kirchler, M., Nave, G., Nosek, B.A., Pfeiffer, T., Altmejd, A., Buttrick, N., Chan, T., Chen, Y., Forsell, E., Gampa, A., Heikensten, E., Hummer, L., Imai, T., … Wu, H. (2018). Evaluating the replicability of social science experiments in Nature and Science between 2010 and 2015. *Nature Human Behavior, 2*(9), 637–644. https://doi.org/10.1038/s41562-018-0399-z

Capaldi, E. J., & Proctor, R. W. (2013). Postmodernism and the development of the psychology of science. In G. J. Feist & M. E. Gorman (Eds.), *Handbook of the psychology of science* (pp. 331–352). Springer.

Carapezza, M., & D'Agostino, M. (2010). Logic and the myth of the perfect language. *Logic & Philosophy of Science, 8*(1), 1–29. https://doi.org/10.1093/oso/9780190869816.003.0001

Carpenter, S. K., & DeLosh, E. L. (2005). Application of the testing and spacing effects to name learning. *Applied Cognitive Psychology: The Official Journal of the Society for Applied Research in Memory and Cognition, 19*(5), 619–636. https://doi.org/10.1002/acp.1101

Cartwright, N., & Montuschi, E. (Eds.). (2014). *Philosophy of social science. A new introduction*. Oxford University Press.

Cavallo, A. M. L. (1996). Meaningful learning, reasoning ability, and students' understanding and problem solving of topics in genetics. *Journal of Research in Science Teaching, 33*(6), 625–656. doi.org/10.1002/(SICI)1098-2736(199608)33:6<625::AID-TEA3>3.0.CO;2-Q

Chafetz, J. S., & Fox, M. F. (2006). Gender, hierarchy, and science. In J. S. Chafetz (Ed.), *Handbook of the Sociology of Gender* (pp. 441–457). Springer. https://doi.org/10.1007/0-387-36218-5

Chalmers, D. (2005). Phenomenal concepts and the explanatory gap. In T. Alter & S. Walter (Eds.), *Phenomenal concepts and phenomenal knowledge. New essays on consciousness and physicalism* (pp. 167–194). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195171655.001.0001

Chen, H., & Yang, J. (2020). Multiple exposures enhance both item memory and contextual memory over time. *Frontiers in Psychology, 11*, 565169. https://doi.org/10.3389/fpsyg.2020.565169

Claus, D. B. (1983). *Toward the soul. An inquiry into the meaning of ψυχή before Plato*. Yale University Press.

Cohen, H. F. (2015). *The rise of modern science explained. A comparative history*. Cambridge University Press. https://doi.org/10.1017/CBO9781316340851

Cook, D. A., Levinson, A. J., & Garside, S. (2010). Time and learning efficiency in internet-based learning: a systematic review and meta-analysis. *Advances in Health Sciences Education, 15*, 755–770. https://doi.org/10.1007/s10459-010-9231-x

Coolican, H. (2019). *Research methods and statistics in psychology*. Routledge. https://doi.org/10.4324/9781315201009

Craver, C., & Tabery, J. (2023). Mechanisms in science. In E. N. Zalta & U. Nodelman (Eds.), *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/science-mechanisms/

De Jonge, M., Tabbers, H. K., Pecher, D., & Zeelenberg, R. (2012). The effect of study time distribution on learning and retention: A Goldilocks principle for presentation rate. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 38*(2), 405–412. https://doi.org/10.1037/a0025897

Devitt, M., & Hanley, R. (2006). *The Blackwell guide to the philosophy of language*. Blackwell. https://doi.org/10.1002/9780470757031

Ebbinghaus, H. (1885). *Über das Gedächtnis. Untersuchungen zur experimentellen Psychologie*. Duncker & Humblot. [*Memory: A contribution to experimental psychology*. Translated by H. A. Ruger, & C. E. Bussenius, (1913). Teachers College Press. https://doi.org/10.1037/10011-000]

Elias, N., Martins, H., & Whitley, R. P. (Eds.). (1982). *Scientific establishments and hierarchies*. D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-7729-7

Endress, A. D., & Potter, M. C. (2014). Large capacity temporary visual memory. *Journal of Experimental Psychology: General, 143*(2), 548–565. https://doi.org/10.1037/a0033934

Farrell, R. P. (2003). *Feyerabend and scientific values. Tightrope-walking rationality*. Springer. https://doi.org/10.1007/978-94-017-1542-3

Faul, F., Erdfelder, E., Lang, A. G., & Buchner, A. (2007). G*Power 3: A flexible statistical power analysis program for the social, behavioral, and biomedical sciences. *Behavior Research Methods 39*(2), 175–191. https://doi.org/10.3758/bf03193146

Feyerabend, P. (1975). *Against method. Outline of an anarchistic theory of knowledge*. New Left Books.

Fiedler, K. (2017). What constitutes strong psychological science? The (neglected) role of diagnosticity and a priori theorizing. *Perspectives on Psychological Science, 12*(1), 46–61. https://doi.org/10.1177/1745691616654458

Fiedler, K. (2018). The creative cycle and the growth of psychological science. *Perspectives on Psychological Science, 13*(4), 433–438. https://doi.org/10.1177/174569161774565

Fischborn, M. (2016). Libet-style experiments, neuroscience, and libertarian free will. *Philosophical Psychology, 29(4),* 494–502. https://doi.org/10.1080/09515089.2016.1141399

Fleck, L. (1935/1980). *Entstehung und Entwicklung einer wissenschaftlichen Tatsache. Einführung in die Lehre vom Denkstil und Denkkollektiv*. Suhrkamp [*Genesis and development of a scientific fact.* Translated by F. Bradley & T. J. Trenn, (1979). The University of Chicago Press.]

Franklin, A. (1986). *The neglect of experiment*. Cambridge University Press. https://doi.org/10.1017/CBO9780511624896

Franklin, A. (1990). *Experiment. Right or wrong*. Cambridge University Press. https://doi.org/10.1017/CBO9780511527302

Franklin, A., & Perovic, S. (2023). Experiment in physics. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/archives/fall2023/entries/physics-experiment

Fredrick, W. C., & Walberg, H. J. (1980). Learning as a function of time. *The Journal of Educational Research, 73*(4), 183–194. https://doi.org/10.1080/00220671.1980.10885233

Galison, P. (1987) *How experiments end*. The University of Chicago Press.

Garber, I. (2019). Trends in contemporary psychology. *Journal of Russian & East European Psychology, 56*(5-6), 271–295. https://doi.org/10.1080/10610405.2019.1659698

García-Rodicio, H. (2023). Relevance instructions and memory for text. The role of instruction specificity and text length. *Culture and Education. 35*(1), 94–119. https://doi.org/10.1080/11356405.2022.2121132

Gazzaniga, M. S. (2011). *Who's in charge? Free will and the science of the brain*. Ecco. https://doi.org/10.1080/02604027.2012.762205

Geiselman, R. E. (1977). Memory for prose as a function of learning strategy and inspection time. *Journal of Educational Psychology, 69*(5), 547–555. https://doi.org/10.1037/0022-0663.69.5.547

Gemignani, M., & Peña, E. (2007). Postmodern conceptualizations of culture in social constructionism and cultural studies. *Journal of Theoretical and Philosophical Psychology, 27-28*(2-1), 276–300. https://doi.org/10.1037/h0091297

Gergen, K. J. (2001). Psychological science in a postmodern context. *American Psychologist, 56*(10), 803–813. https://doi.org/10.1037/0003-066X.56.10.803

Gergen, K. J. (2002). Psychological science: to conserve or create? *American Psychologist, 57*(6-7), 463–464. https://doi.org/10.1037/0003-066X.57.6-7.463

Gerrig, R. J. (2012). *Psychology and life* (20th ed.). Pearson.

Gettinger, M. (1991). Learning time and retention differences between nondisabled students and students with learning disabilities. *Learning Disability Quarterly, 14*(3), 179–189. https://doi.org/10.2307/1510848

Glennan, S. S. (1996). Mechanisms and the nature of causation. *Erkenntnis, 44*, 49–71. https://doi.org/10.1007/BF00172853

Glennan, S. (2002). Rethinking mechanistic explanation. *Philosophy of Science, 69*(S3), 342–353. https://doi.org/10.1086/341857

Godfrey-Smith, P. (2021). *Theory and reality. An introduction to the philosophy of science* (2nd ed.). The University of Chicago Press.

Goldstein, E. B. (2015). *Cognitive psychology: Connecting mind, research and everyday experience*. Cengage Learning.

Goldthorpe, J. H. (2001). Causation, statistics, and sociology. *European Sociological Review, 17*(1), 1–20. https://doi.org/10.1093/esr/17.1.1

Goodwin, C. J. (2015). *A history of modern psychology* (5th ed.). Wiley.

Gower, B. (1997). *Scientific method. An historical and philosophical introduction*. Routledge. https://doi.org/10.4324/9780203046128

Greenwood, J. D. (2009). *A conceptual history of psychology* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9781107414914

Haaga, D. A. (2004). Defining psychology: what can it do for us? *Journal of Clinical Psychology, 60*(12), 1227–1229. https://doi.org/10.1002/jclp.20063

Hackert, B., & Weger, U. (2018). Introspection and the Würzburg school: implications for experimental psychology today. *European Psychologist, 23*, 217–232. https://doi.org/10.1027/1016-9040/a000329

Haig, B. D. (2014). *Investigating the psychological world: Scientific method in the behavioral sciences*. MIT Press. https://doi.org/10.7551/mitpress/9780262027366.001.0001

Haig, B. D. (2018a). The importance of scientific method for psychological science. *Psychology, Crime & Law, 25*(6), 527–541. https://doi.org/10.1080/1068316X.2018.1557181

Haig, B. D. (2018b). *Method matters in psychology. Essays in applied philosophy of science*. Springer. https://doi.org/10.1007/978-3-030-01051-5

Halamish, V., & Bjork, R. A. (2011). When does testing enhance retention? A distribution-based interpretation of retrieval as a memory modifier. *Journal of Experimental Psychology: Learning, Memory, and Cognition, 37*(4), 801–812. https://doi.org/10.1037/a0023219

Hansen, J. T. (2015). The relevance of postmodernism to counselors and counselling practice. *Journal of Mental Health Counseling, 37*(4), 355–363. https://doi.org/10.17744/MEHC.37.4.06

Hare-Mustin, R. T., & Marecek, J. (1988). The meaning of difference: Gender theory, postmodernism, and psychology. *American Psychologist, 43*(6), 455–464. https://doi.org/10.1037//0003-066X.43.6.455

Harris, R. (2005). *The semantics of science*. Continuum.

Hartwig, M. K., & Dunlosky, J. (2012). Study strategies of college students: Are self-testing and scheduling related to achievement? *Psychonomic Bulletin & Review, 19*, 126–134. https://doi.org/10.3758/s13423-011-0181-y

Hearnshaw, L. S. (1941). Psychology and operationism. *Australasian Journal of Psychology and Philosophy, 19*(1), 44–57. https://doi.org/10.1080/00048404108541506

Heidegger, M. (1953). *Sein und Zeit* (7. Aufl.). [*Being and Time*. Translated by J. Stambaugh, revised by D. J. Schmidt, (2010). State University of New York Press.]

Held, B. S. (2007). *Psychology's interpretive turn. The search for truth and agency in theoretical and philosophical psychology*. American Psychological Association. https://doi.org/10.1037/11588-000

Henriques, G. R. (2004). Psychology defined. *Journal of Clinical Psychology, 60*(12), 1207–1221. https://doi.org/10.1002/jclp.20061

Henry, J. (1997). *The scientific revolution and the origins of modern science*. Macmillan. https://doi.org/10.1007/978-1-349-25512-2

Hibberd, F. J. (2019). What is scientific definition? *The Journal of Mind and Behavior, 40*(1), 29–52.

Hodge, B. (1999). The Sokal 'Hoax': some implications for science and postmodernism. *Continuum: Journal of Media & Cultural Studies, 13*(2), 255–269. https://doi.org/10.1080/10304319909365797

Hofmann, S. G. (2002). More science, not less. *American Psychologist, 57*(6–7), 462. https://doi.org/10.1037//0003-066X.57.6-7.462a

Hogan, T. P. (2019). *Psychological testing: A practical introduction* (4th ed.). Wiley.

Hollis, M. (1994). *The philosophy of social science. An introduction*. Cambridge University Press. https://doi.org/10.1017/CCOL0521447801

Holt, R. R. (2005). The menace of postmodernism to a psychoanalytic psychology. I J. S. Auerbach, K. N. Levy, & C. E. Schaffer (Eds.), *Relatedness, self-definition and mental representation. Essays in honor of Sidney J. Blatt* (pp. 288–302). Routledge. https://doi.org/10.4324/9780203337318_chapter_18

Holtz, P. (2020). Does postmodernism really entail a disregard for the truth? Similarities and differences in postmodern and critical rationalist conceptualizations of truth, progress, and empirical research methods. *Frontiers in Psychology, 11,* 545959. https://doi.org/10.3389/fpsyg.2020.545959

Holzman, L. (2006). Activating postmodernism. *Theory & Psychology, 16*(1), 109–123. https://doi.org/10.1177/0959354306060110

Holzman, L., & Morss, J. (Eds). (2000). *Postmodern psychologies, societal practice, and political life*. Routledge. https://doi.org/10.4324/9781315811314

Hothersall, D. (2004). *History of Psychology* (5th ed.). Cambridge University Press. https://doi.org/10.1017/9781108774567

Howell, D. C. (2010). *Statistical methods for psychology* (7th ed.). Wadsworth.

Hutmacher, F. (2019a). On the janus-facedness of stress and modern life. *Journal of Theoretical and Philosophical Psychology, 39*(3), 181–192. https://doi.org/10.1037/teo0000113

Hutmacher, F. (2019b). Why is there so much more research on vision than on any other sensory modality? *Frontiers in Psychology, 10*, 2246. https://doi.org/10.3389/fpsyg.2019.02246

Hutmacher, F., & Mayrhofer, R. (2023). Psychology as a historical science? Theoretical assumptions, methodological considerations, and potential pitfalls. *Current Psychology, 42*(22), 18507–18514. https://doi.org/10.1007/s12144-022-03030-0

Jager, B. (1991). Psychology in a postmodern era. *Journal of Phenomenological Psychology, 22*(1), 60–71. https://doi.org/10.1163/156916291X00046

Jagtenberg, T. (1983). *The social construction of science. A comparative study of goal direction, research evolution and legitimation*. D. Reidel Publishing Company. https://doi.org/10.1007/978-94-009-7010-6

Jiménez, J. P. (2015). Psychoanalysis in postmodern times: some questions and challenges. *Psychoanalytic Inquiry, 35*(6), 609–624. https://doi.org/10.1080/07351690.2015.1055221

Jonassen, D. H., & Strobel, J. (2006). Modeling for meaningful learning. In D. Hung & M. S. Khine (Eds.). *Engaged learning with emerging technologies* (pp. 1–27). Springer. https://doi.org/10.1007/1-4020-3669-8_1

Jüttemann, G. (1983). Psychologie am Scheideweg: Teilung oder Vervollständigung?, in G. Jüttemann (Ed.), *Psychologie in der Veränderung. Perspektiven für eine gegenstandsangemessenere Forschungspraxis* (pp. 30–65). Beltz.

Kantorovich, A. (1988). Philosophy of science: From justification to explanation. *The British Journal for the Philosophy of Science, 39*(4), 469–494. https://doi.org/10.1093/bjps/39.4.469

Karpicke, J. D., & Blunt, J. R. (2011). Retrieval practice produces more learning than elaborative studying with concept mapping. *Science, 331*(6018), 772–775. https://doi.org/10.1126/science.1199327

Kauffman, S. A. (1970). Articulation of parts explanation in biology and the rational search for them. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1970*, 257–272. https://doi.org/10.1086/psaprocbienmeetp.1970.495768

Keuth, H. (2005). *The philosophy of Karl Popper*. Cambridge University Press.

Kihlstrom, J. F. (2017). Time to lay the Libet experiment to rest: commentary on Papanicolaou (2017). *Psychology of Consciousness: Theory, Research, and Practice, 4*(3), 324–329. https://doi.org/10.1037/cns0000124

Kim, A. S. N., Wong-Kee-You, A. M. B., Wiseheart, M., & Rosenbaum, R. S. (2019). The spacing effect stands up to big data. *Behavior Research Methods, 51*, 1485–1497. https://doi.org/10.3758/s13428-018-1184-7

Kincaid, H. (Ed.). (2012). *The Oxford handbook of philosophy of social science*. Oxford University Press. https://doi.org/10.1093/oxfordhb/9780195392753.001.0001

King, D. B., Woody, W. D., & Viney, W. (2013). *A history of psychology. Ideas & context.* Pearson.

Kleinert, A. (2009). Der messende Luchs. *NTM. Zeitschrift für Geschichte der Wissenschaften, Technik und Medizin, 17*, 199–206. https://doi.org/10.1007/s00048-009-0335-4

Kornell, N., Bjork, R. A., & Garcia, M. A. (2011). Why tests appear to prevent forgetting: A distribution-based bifurcation model. *Journal of Memory and Language, 65*(2), 85–97. https://doi.org/10.1016/j.jml.2011.04.002

Krippner, S. (2001). Research methodology in humanistic psychology in the light of postmodernity. In K. J. Schneider, J. F. Bugental, & J. F. Pierson (Eds.). *The handbook of humanistic psychology: Leading edges in theory, research, and practice* (pp. 290–304). SAGE. https://doi.org/10.4135/9781412976268.n22

Kruger, D. J. (2002). The deconstruction of constructivism. *American Psychologist, 57*(6–7), 456–457. https://doi.org/10.1037/0003-066X.57.6-7.456

Kuhbandner, C., & Emmerdinger, K. J. (2019). Do students really prefer repeated rereading over testing when studying textbooks? A reexamination. *Memory 27*(7), 952–961. https://doi.org/10.1080/09658211.2019.1610177

Kuhn, T. S. (1962). *The structure of scientific revolutions*. The University of Chicago Press.

Kvale, S. (ed.). (1992). *Psychology and postmodernism*. SAGE.

Lakatos, I. (1978). *The methodology of scientific research programmes*. Cambridge University Press. https://doi.org/10.1017/CBO9780511621123

Langdridge, D. (2007). *Phenomenological psychology. Theory, research and method*. Pearson.

Latour, B. (1999). *Pandora's hope. Essays on the reality of science studies*. Harvard University Press.

Lavazza, A. (2016). Free will and neuroscience: from explaining freedom away to new ways of operationalizing and measuring it. *Frontiers in Human Neuroscience, 10*, 262. https://doi.org/10.3389/fnhum.2016.00262

Law, J. (2004). *After method: Mess in social science research*. Routledge. https://doi.org/10.4324/9780203481141

Leahey, T. H. (2018). *A history of psychology. From antiquity to modernity* (8th ed.). Routledge, Taylor & Francis. https://doi.org/10.4324/9781315624273

Lechuga, M. T., Ortega-Tudela, J. M., & Gómez-Ariza, C. J. (2015). Further evidence that concept mapping is not better than repeated retrieval as a tool for learning from texts. *Learning and Instruction, 40*, 61–68. https://doi.org/10.1016/j.learninstruc.2015.08.002

Leezenberg, M., & De Vries, G. (2018). *History and philosophy of the humanities. An introduction*. Amsterdam University Press. https://doi.org/10.2307/j.ctvnp0j8d

Leffert, M. (2007). A contemporary integration of modern and postmodern trends in psychoanalysis. *Journal of the American Psychoanalytic Association, 55*(1), 177–197. https://doi.org/10.1177/00030651070550011001

Libet, B., Gleason, C. A., Wright, E. W., & Pearl, D. K. (1983). Time of conscious intention to act in relation to onset of cerebral activity (readiness-potential). The unconscious initiation of a freely voluntary act. *Brain 106*, 623–642. https://doi.org/10.1093/brain/106.3.623

Lilienfeld, S. O. (2004). Defining psychology: is it worth the trouble?. *Journal of Clinical Psychology, 60*(12), 1249–1253. https://doi.org/10.1002/jclp.20067

Lilienfeld, S. O. (2010). Can psychology become a science?. *Personality and Individual Differences, 49*(4), 281–288. https://doi.org/10.1016/j.paid.2010.01.024

Lilienfeld, S. O., Lynn, S. J., Namy, L., & Woolf, N. (2015). *Psychology. From inquiry to understanding*. Pearson.

Lincoln, J., & Hoffman, L. (2019). Toward an integration of constructivism and existential psychotherapy. *Journal of Constructivist Psychology. 32*(2) 108–125. https://doi.org/10.1080/10720537.2018.1461719

Locke, E. A. (2002). The dead end of postmodernism. *American Psychologist, 57*(6–7), 458. https://doi.org/10.1037/0003-066X.57.6-7.458a

Luck, S. J., & Vogel, E. K. (1997). The capacity of visual working memory for features and conjunctions. *Nature, 390*(6657), 279–281. https://doi.org/10.1038/36846

Luck, S. J., & Vogel, E. K. (2013). Visual working memory capacity: from psychophysics and neurobiology to individual differences. *Trends in Cognitive Sciences, 17*(8), 391–400. https://doi.org/10.1016/j.tics.2013.06.006

Lunt, I. (1999). The professionalization of psychology in Europe. *European Psychologist, 4*(4), 240–247. https://doi.org/10.1027/1016-9040.4.4.240

Lycan, W. G. (2019). *Philosophy of language. A contemporary introduction*. Routledge.

Ma, W. J., Husain, M., & Bays, P. M. (2014). Changing concepts of working memory. *Nature Neuroscience, 17*(3), 347–356. https://doi.org/10.1038/nn.3655

Machamer, P., & Silberstein, M. (Eds.). (2002). *The Blackwell guide to the philosophy of science*. Blackwell. https://doi.org/10.1002/9780470756614

Malich, L., & Rehmann-Sutter, C. (2022). Metascience is not enough – a plea for psychological humanities in the wake of the replication crisis. *Review of General Psychology, 26*(2), 261–273. https://doi.org/10.1177/10892680221083876

Mandler, G. (2007). *A history of modern experimental psychology. From James and Wundt to cognitive science*. The MIT Press. https://doi.org/10.7551/mitpress/3542.001.0001

Margolis, J. (1995). *Historied thought, constructed world. A conceptual primer for the turn of the millennium*. University of California Press.

Margolis, J. (2008). *The arts and the definition of the human. Toward a philosophical anthropology*. Stanford University Press. https://doi.org/10.1515/9780804769860

Martin, J. (2003). Positivism, quantification and the phenomena of psychology. *Theory & Psychology, 13*(1), 33–38. https://doi.org/10.1177/0959354303013001760

Martin, J., & Sugarman, J. (2009). Middle-ground theorizing, realism, and objectivity in psychology: a commentary on Held (2007). *Theory & Psychology, 19*(1), 115–122. https://doi.org/10.1177/0959354308101422

Matlin, M. W. (2012). *Cognition* (8th ed.). Wiley.

Matthew, D. (2005). *Science in society*. Palgrave Macmillan.

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*(4), 226–232. https://doi.org/10.1207/s15430421tip4104_4

Mayrhofer, R., & Hutmacher, F. (2020). The Principle of inversion: Why the quantitative-empirical paradigm cannot serve as a unifying basis for psychology as an academic discipline. *Frontiers in Psychology, 11*, 596425. https://doi.org/10.3389/fpsyg.2020.596425

Mayrhofer, R., Kuhbandner, C., & Frischholz, K. (2023). Re-examining the testing effect as a learning strategy: the advantage of retrieval practice over concept mapping as a methodological artifact. *Frontiers in Psychology, 14*, 1258359. https://doi.org/10.3389/fpsyg.2023.1258359

Mayrhofer, R., Kuhbandner, C., & Lindner, C. (2021). The practice of experimental psychology. An inevitably postmodern endeavor. *Frontiers in Psychology, 11*, 612805. https://doi.org/10.3389/fpsyg.2020.612805

McCrudden, M. T., Schraw, G., & Kambe, G. (2005). The effect of relevance instructions on reading time and learning. *Journal of Educational Psychology, 97*(1), 88–102. https://doi.org/10.1037/0022-0663.97.1.88

McDonald, R. P. (2013). *Test theory: A unified treatment*. Psychology Press. https://doi.org/10.4324/9781410601087

McKenna, M., & Pereboom, D. (2016). *Free will. A contemporary introduction*. Routledge, Taylor & Francis. https://doi.org/10.4324/9781315621548

Mele, A. R. (2007). Decisions, intentions, urges, and free will. Why Libet has not shown what he says he has. In J. K. Campbell, M. O'Rourke, & H. S. Silverstein (Eds.), *Causation and Explanation* (pp. 241–263). MIT Press.

Michell, J. (1997). Quantitative science and the definition of measurement in psychology. *British Journal of Psychology, 88*(3), 355–383. https://doi.org/10.1111/j.2044-8295.1997.tb02641.x

Michell, J. (2003). The quantitative imperative: Positivism, naïve realism and the place of qualitative methods in psychology. *Theory & Psychology, 13*(1), 5–31. https://doi.org/10.1177/0959354303013001758

Mintzes, J. J., Canas, A., Coffey, J., Gorman, J., Gurley, L., Hoffman, R., McGuire, S. Y., Miller, N., Moon, B., Trifone, J., & Wandersee, J.H. (2011). Comment on "retrieval practice produces more learning than elaborative studying with concept mapping". *Science 334*(6055), 453. https://doi.org/10.1126/science.1203698

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

Mulkay, M. J. (1970). Conformity and innovation in science. *The Sociological Review, 18*(1), 5–23. https://doi.org/10.1111/j.1467-954X.1970.tb03173.x

Mulkay, M. J. (1976). Norms and ideology in science. *Social Science Information, 15*(4–5), 637–656. https://doi.org/10.1177/053901847601500406

Murdock, B. B. (1960). The immediate retention of unrelated words. *Journal of Experimental Psychology, 60*(4), 222–234. https://doi.org/10.1037/h0045145

Murphy, D. H., Bjork, R. A., & Bjork, E. L. (2022). Going beyond the spacing effect: Does it matter how time on a task is distributed? *Quarterly Journal of Experimental Psychology, 76*(5), 1131–1153. https://doi.org/10.1177/17470218221113933

Myers, D. G., & Dewall, C. N. (2015). *Psychology* (11th ed.) Worth Publishers.

Nahmias, E. (2010). Scientific challenges to free will. In C. Sandis & T. O'Connor (Eds.), *A companion to the philosophy of action* (pp. 345–356). Wiley-Blackwell. https://doi.org/10.1002/9781444323528.ch44

Niaz, M. (2020). *Feyerabend's epistemological anarchism. How science works and its importance for science education*. Springer. https://doi.org/10.1007/978-3-030-36859-3

Nickerson, R. (1998). Confirmation bias: A ubiquitous phenomenon in many guises. *Review of General Psychology, 2*(2), 175–220. https://doi.org/10.1037/1089-2680.2.2.175

Nola, R., & Sankey, H. (2014). *Theories of scientific method. An introduction*. Acumen. https://doi.org/10.4324/9781315711959

Novak, J. D. (1995). Concept mapping to facilitate teaching and learning. *Prospects, 25*(1), 79–86. https://doi.org/10.1007/BF02334286

Novak, J. D., & Cañas, A. J. (2006). The origins of the concept mapping tool and the continuing evolution of the tool. *Information Visualization, 5*(3), 175–184. https://doi.org/10.1057/palgrave.ivs.9500126

Novick, M. R. (1966). The axioms and principal results of classical test theory. *Journal of Mathematical Psychology, 3*(1), 1–18. https://doi.org/10.1016/0022-2496(66)90002-2

Oberheim, E. (2006). *Feyerabend's philosophy*. De Gruyter. https://doi.org/10.1515/9783110891768

O'Connor, T., & Franklin, C. (2018). Free will. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy.* https://plato.stanford.edu/entries/freewill/

O'Day, G. M., & Karpicke, J. D. (2021). Comparing and combining retrieval practice and concept mapping. *Journal of Educational Psychology, 113*(5), 986–997. https://doi.org/10.1037/edu0000486

Olby, R. C., Cantor, G. N., Christie, J. R. R., & Hodge, M. J. S. (Eds). (1991). *Companion to the history of modern science*. Routledge. https://doi.org/10.4324/9781003070818

Osterlind, S. J. (2019). *The error of truth: How history and mathematics came together to form our character and shape our worldview*. Oxford University Press. https://doi.org/10.1093/oso/9780198831600.001.0001

Parkin, A. J. (2013). *Essential cognitive psychology* (classic ed.). Psychology Press. https://doi.org/10.4324/9780203724637

Paul, L. A., & Hall, N. (2013). *Causation: A user's guide*. Oxford University Press. https://doi.org/10.1093/acprof:oso/9780199673445.001.0001

Pearl, J. (2009). *Causality. Models, reasoning, and inference* (2nd ed.). Cambridge University Press. https://doi.org/10.1017/CBO9780511803161

Petrie, H. G. (1971). A dogma of operationalism in the social sciences. *Philosophy of the Social Sciences, 1*(1), 145–160. https://doi.org/10.1177/004839317100100109

Pinker, S. (2018). *Enlightenment now. The case for reason, science, humanism, and progress*. Viking.

Planck, M. (1949). The meaning and limits of exact science. *Science, 110*(2857), 319–327. https://doi.org/10.1126/science.110.2857.319

Platts, M. (1997). *Ways of meaning. An introduction to philosophy of language* (2nd ed.). MIT Press.

Preston, J. (1997). *Feyerabend: philosophy, science and society*. Polity Press.

Psillos, S. (2007). *Philosophy of science A-Z*. Edinburgh University Press. https://doi.org/10.1515/9780748629145

Radder, H., & Meynen, G. (2013). Does the brain "initiate" freely willed processes? A philosophy of science critique of Libet-type experiments and their interpretation. *Theory & Psychology, 23*(1), 3–21. https://doi.org/10.1177/0959354312460926

Ramey, H. L., & Grubb, S. (2009). Modernism, postmodernism and (evidence-based) practice. *Contemporary Family Therapy, 31*, 75–86. https://doi.org/10.1007/s10591-009-9086-6

Reber, R. (2017). Availability. In R. F. Pohl (Ed.), *Cognitive illusions. Intriguing phenomena in thinking, judgment and memory* (2nd ed.). Routledge. https://doi.org/10.4324/9781315696935

Richardson, A., & Uebel, T. (Eds.). (2007). *The Cambridge companion to logical empiricism*. Cambridge University Press. https://doi.org/10.1017/CCOL0521791782

Robins, S., Symons, J., & Calvo, P. (Eds.). (2020). *The Routledge companion to philosophy of psychology* (2nd ed.). Routledge. https://doi.org/10.4324/9780429244629

Robinson, M. M., Benjamin, A. S., & Irwin, D. E. (2020). Is there a K in capacity? Assessing the structure of visual short-term memory. *Cognitive Psychology, 121*, 101305. https://doi.org/10.1016/j.cogpsych.2020.101305

Roediger, H. L., & Karpicke, J. D. (2006). Test-enhanced learning: Taking memory tests improves long-term retention. *Psychological Science, 17*(3), 249–255. https://doi.org/10.1111/j.1467-9280.2006.01693.x

Roelle, J., Lehmkuhl, N., Beyer, M. U., & Berthold, K. (2015). The role of specificity, targeted learning activities, and prior knowledge for the effects of relevance instructions. *Journal of Educational Psychology, 107*(3), 705–723. https://doi.org/10.1037/edu0000010

Rohrer, D., & Pashler, H. (2007). Increasing retention without increasing study time. *Current Directions in Psychological Science, 16*(4), 183–186. https://doi.org/10.1111/j.1467-8721.2007.00500.x

Ronzitti, G. (Ed.). (2011). *Vagueness: A guide*. Springer. https://doi.org/10.1007/978-94-007-0375-9

Roskies, A. L. (2011). Why Libet's studies don't pose a threat to free will. In W. Sinnott-Armstrong & L. Nadel (Eds.), *Conscious will and responsibility* (pp. 11–22). Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195381641.003.0003

Rowbottom, D. P. (2011). *Popper's critical rationalism. A philosophical investigation*. Routledge. https://doi.org/10.4324/9780203836187

Rowland, C. A. (2014). The effect of testing versus restudy on retention: a metanalytic review of the testing effect. *Psychological Bulletin, 140*(6), 1432–1463. https://doi.org/10.1037/a0037559

Royce, J. R. (1961). Heretical thoughts on the definition of psychology. *Psychological Reports, 8*(1), 11–14. https://doi.org/10.2466/pr0.1961.8.1.11

Russo Krauss, C. (2019). The spread of the new definition of psychology. In C. Russo Krauss (Ed.), *Wundt, Avenarius, and scientific psychology* (pp. 59–111). Palgrave Macmillan. https://doi.org/10.1007/978-3-030-12637-7_4

Saigle, V., Dubljević, V., & Racine, E. (2018). The impact of a landmark neuroscience study on free will: a qualitative analysis of articles using Libet and colleagues' methods. *AJOB Neuroscience, 9*(1), 29–41. https://doi.org/10.1080/21507740.2018.1425756

Sankey, H. (2014). Scientific method. In M. Curd & S. Psillos (Eds.), *The Routledge companion to philosophy of science* (pp. 280–290). Routledge. https://doi.org/10.4324/9780203744857

Scheel, A. M. (2022). Why most psychological research findings are not even wrong. *Infant and Child Development, 31*(1), e2295. https://doi.org/10.1002/icd.2295

Scheel, A. M., Tiokhin, L., Isager, P. M., & Lakens, D. (2021). Why hypothesis testers should spend less time testing hypotheses. *Perspectives on Psychological Science, 16*(4), 744–755. https://doi.org/10.1177/1745691620966795

Schlosser, M. E. (2014). The neuroscientific study of free will: a diagnosis of the controversy. *Synthese 191*(2), 245–262. https://doi.org/10.1007/s11229-013-0312-2

Schurger, A. (2017). The neuropsychology of conscious volition. In S. Schneider & M. Velmans (Eds.), *The Blackwell companion to consciousness* (pp. 695–710). Wiley Blackwell. https://doi.org/10.1002/9781119132363.ch49

Schurgin, M. W. (2018). Visual memory, the long and the short of it: a review of visual working memory and long-term memory. *Attention, Perception, & Psychophysics, 80*, 1035–1056. https://doi.org/10.3758/s13414-018-1522-y

Schurgin, M. W., & Brady, T. F. (2019). When "capacity" changes with set size: ensemble representations support the detection of across-category changes in visual working memory. *Journal of Vision, 19*(5), 1–13. https://doi.org/10.1167/19.5.3

Schurgin, M. W., Cunningham, C. A., Egeth, H. E., & Brady, T. F. (2018). Visual long-term memory can replace active maintenance in visual working memory. *bioRxiv* [Preprint]. https://doi.org/10.1101/381848

Schwieren, J., Barenberg, J., & Dutke, S. (2017). The testing effect in the psychology classroom: a meta-analytic perspective. *Psychology Learning & Teaching, 16*(2), 179–196. https://doi.org/10.1177/1475725717695149

Segerstråle, U. C. O. (Ed.). (2000). *Beyond the science wars: The missing discourse about science and society*. State University of New York Press.

Seifert, J. (2011). In defense of free will: a critique of Benjamin Libet. *The Review of Metaphysiscs. 65*(2), 377–407.

Shaughnessy, J. J., Zechmeister, E. B., & Zechmeister, J. S. (2015). *Research methods in psychology* (10th ed.). McGraw-Hill.

Shearer, A. (2020). *The story of yoga. From ancient India to the modern West*. Hurst & Company.

Shepard, J. (2018). How Libet-style experiments may (or may not) challenge lay theories of free will. *AJOB Neuroscience, 9*(1), 45–47. https://doi.org/10.1080/21507740.2018.1425766

Shibayama, S., & Baba, Y. (2015). Dishonest conformity in peer review. *Prometheus, 33*(3), 215–233. https://doi.org/10.1080/08109028.2015.1114745

Shuell, T. J. (1990). Phases of meaningful learning. *Review of educational research, 60*(4), 531–547. https://doi.org/10.3102/00346543060004531

Sim, S. (2011). *The Routledge companion to postmodernism* (3rd ed.). Routledge. https://doi.org/10.4324/9780203813201

Sinacore, A. L., & Enns, C. Z. (2005). Diversity feminisms: postmodern, women-of-color, antiracist, lesbian, third-wave, and global perspectives. In C. Z. Enns & A. L. Sinacore (Eds.), *Teaching and social justice. Integrating multicultural and feminist theories in the classroom* (pp. 41–68). American Psychological Association. https://doi.org/10.1037/10929-003

Slavin, R. E. (2002). Evidence-based education policies: transforming educational practice and research. *Educational Researcher, 31*(7), 15–21. https://doi.org/10.3102/0013189X031007015

Smedslund, J. (2016). Why psychology cannot be an empirical science. *Integrative Psychological and Behavioral Science, 50*, 185–195. https://doi.org/10.1007/s12124-015-9339-x

Sokal, A. D. (1996). A physicist experiments with cultural studies. *Lingua Franca 6*(4), 62–64.

Staley, K.W. (2014). *An introduction to the philosophy of science*. Cambridge University Press. https://doi.org/10.1017/CBO9781139047760

Sternberg, R. J., & Sternberg, K. (2011). *Cognitive psychology* (6th ed.). Cengage Learning.

Stroebe, W., & Strack, F. (2014). The alleged crisis and the illusion of exact replication. *Perspectives on Psychological Science, 9*(1), 59–71. https://doi.org/10.1177/1745691613514450

Summers, F. (2012). Psychoanalysis, the tyranny of objectivism, and the rebellion of the subjective. *International Journal of Applied Psychoanalytic Studies, 9*(1), 35–47. https://doi.org/10.1002/aps.310

Szostak, R. (2004). *Classifying science. Phenomena, data, theory, method, practice*. Springer. https://doi.org/10.1007/978-1-4020-3095-6

Taleb, N. N. (2007). *The black swan. The impact of the highly improbable*. Random House.

Tarescavage, A. M. (2020). Science wars II: the insidious influence of postmodern ideology on clinical psychology (commentary on "Implications of ideological bias in social psychology on clinical practice"). Clinical Psychology: *Science and Practice, 27*(2), e12319. https://doi.org/10.1111/cpsp.12319

Teo, T. (2018). *Outline of theoretical psychology. Critical investigations*. Palgrave Macmillan. https://doi.org/10.1057/978-1-137-59651-2

Thagard, P. (Ed.). (2007). *Philosophy of psychology and cognitive science*. North-Holland.

Thomas, E. L., & Robinson, H. A. (1972). *Improving reading in every class: A sourcebook for teachers* (3rd ed.). Allyn and Bacon.

Toomela, A. (2020). Psychology today: still in denial, still outdated. *Integrative Psychological and Behavioral Science, 54*, 563–571. https://doi.org/10.1007/s12124-020-09534-3

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology, 5*(2), 207–232. https://doi.org/10.1016/0010-0285(73)90033-9

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 185*(4175), 1124–1131. https://doi.org/10.1126/science.185.4157.11

Uebel, T., & Limbeck-Lilienau, C. (Eds.). (2022). *The Routledge handbook of logical empiricism*. Routledge. https://doi.org/10.4324/9781315650647

Vallori, A. B. (2014). Meaningful learning in practice. *Journal of Education and Human Development, 3*(4). https://doi.org/10.15640/jehd.v3n4a18

Valsiner, J. (2012). *A guided science. History of psychology in the mirror of its making*. Routledge. https://doi.org/10.4324/9781315083544

VandenBos, G. R. (Ed.). (2015). *APA dictionary of psychology* (2nd ed.). American Psychological Association. https://doi.org/10.1037/14646-000

Van Fraassen, B. (1980). *The scientific image*. Oxford University Press. https://doi.org/10.1093/0198244274.001.0001

Walsh, R. T. G., Teo, T., & Baydala, A. (2014). *A critical history and philosophy of psychology*. Cambridge University Press. https://doi.org/10.1017/CBO9781139046831

Warmoth, A. (1998). Humanistic psychology and humanistic social science. *Humanity & Society, 22*(3), 313–319. https://doi.org/10.1177/016059769802200306

Wason, P. C. (1960). On the failure to eliminate hypotheses in a conceptual task. *Quarterly journal of experimental psychology, 12*(3), 129–140. https://doi.org/10.1080/17470216008416

Wason, P. C. (1968). Reasoning about a rule. *Quarterly journal of experimental psychology, 20*(3), 273–281. https://doi.org/10.1080/14640746808400161

Watanabe, T. (2010). Metascientific foundations for pluralism in psychology. *New Ideas in Psychology, 28*(2), 253–262. https://doi.org/10.1016/j.newideapsych.2009.09.019

Wax, M. L. (1995). Method as madness science, hermeneutics, and art in psychoanalysis. *Journal of the American Academy of Psychoanalysis, 23*(4), 525–543. https://doi.org/10.1521/jaap.1.1995.23.4.525

Wegner, D. M. (2017). *The illusion of conscious will* (2nd ed.). MIT Press.

Weiskopf, D. A., & Adams, F. (2015). *An introduction to the philosophy of psychology*. Cambridge University Press. https://doi.org/10.1017/CBO9781139019040

Weissgerber, S. C., & Rummer, R. (2023). More accurate than assumed: Learners' metacognitive beliefs about the effectiveness of retrieval practice. *Learning and Instruction, 83*, 101679. https://doi.org/10.1016/j.learninstruc.2022.101679

Wiggins, B. J., & Christopherson, C. D. (2019). The replication crisis in psychology: An overview for theoretical and philosophical psychology. *Journal of Theoretical and Philosophical Psychology, 39*(4), 202–217. https://doi.org/10.1037/teo0000137

Wilton, R., & Harley, T. (2017). *Science and psychology*. Routledge. https://doi.org/10.4324/9781315529295

Yang, C., Luo, L., Vadillo, M. A., Yu, R., & Shanks, D. R. (2021). Testing (quizzing) boosts classroom learning: A systematic and meta-analytic review. *Psychological Bulletin, 147*(4), 399–435. https://doi.org/10.1037/bul0000309

Yang, J., Zhan, L., Wang, Y., Du, X., Zhou, W., Ning, X., Sun, Q., & Moscovitch, M. (2016). Effects of learning experience on forgetting rates of item and associative memories. *Learning & Memory, 23*(7), 365–378. https://doi.org/10.1101/lm.041210.115

Zacks, R. T. (1969). Invariance of total learning time under different conditions of practice. *Journal of Experimental Psychology, 82*, 441–447. https://doi.org/10.1037/h0028369

Zagaria, A., Ando, A., & Zennaro, A. (2020). Psychology: a giant with feet of clay. *Integrative Psychological and Behavioral Science, 54*, 521–562. https://doi.org/10.1007/s12124-020-09524-5

Zinman, J. (2000). *Real science. What it is, and what it means*. Cambridge University Press. https://doi.org/10.1017/CBO9780511541391