

# Early Experimental Psychology: How did Replication Work Before P-Hacking?

Annette Mülberger<sup>1</sup> 

Review of General Psychology  
2022, Vol. 26(2) 131–145  
© The Author(s) 2022



Article reuse guidelines:

[sagepub.com/journals-permissions](https://sagepub.com/journals-permissions)  
DOI: 10.1177/10892680211066468  
[journals.sagepub.com/home/rgp](https://journals.sagepub.com/home/rgp)



## Keywords

Experimentation, Psychophysics, Repetition, Scientific Practice, Nineteenth Century, Controversy

## Introduction

For many researchers, replication is still the “gold standard” that is crucial for verifying scientific findings (see, for example, Frank & Saxe, 2012; Iso-Ahola, 2020; Witte & Zenker, 2017). Indeed, Crandall and Sherman (2016) declared that: “[t]here is no controversy over the need for replication; virtually all scientists and philosophers of science endorse the notion that replication of one sort or another is absolutely essential” (p. 94). In recent decades, this has led to widespread concern because few experimental findings are actually being confirmed in this way (see, for example, Pashler & Wagenmakers, 2012; Reproducibility Project: Psychology<sup>1</sup>; Wiggins & Chrispherson, 2019).

Before it is possible to plan how to remedy this situation, the reasons for the lack of replications must be identified. “Questionable research practices” such as p-hacking or post-hoc hypothesizing, the “file-drawer problem,”<sup>2</sup> are often cited as contributing to the problem (Romero, 2019; Wiggins & Chrispherson, 2019). These research practices are firmly embedded within a scientific culture that is characterized by a highly competitive academic environment and a reward system that dissuades rather than encouraging replication (Crandall & Sherman, 2016; Romero, 2019). This setting fosters personal ambition, urging researchers to come up with innovative and ambitious projects continually and to publish as many papers as possible. Meanwhile, most journals only publish reports of original research offering statistically significant results, which has led to a “publication bias” (Romero, 2017). Replicability problems, as Pashler and Wagenmakers (2012) stated, “reflect deep-seated human biases and well entrenched incentives that shape the behavior of individuals and institutions” (p. 529). Fraud cases, such as that involving Diederik Stapel, show just how far a person might be willing to go when succumbing to such pressures (Stroebe, Postmes, & Spears, 2012; Derksen, 2021).

Whether replication is really necessary and whether the problematic research practices mentioned above are due more to the present reward system, general human biases or an incorrect statistical or philosophical understanding are still open questions (Feest, 2019; Flis, 2019; Morawski, 2019). Given such uncertainties, it seems interesting to explore how research was

undertaken in the past, when the current institutional conditions did not pertain—or at least, not yet fully. Stated differently: If the current replicability problem is related to recent research practices that have appeared as part of academic life in times of neoliberal capitalism and “big science,” then we might assume that replication worked differently in the past. Thus, in the present paper, I adopt a historical stance to reveal characteristics of nineteenth-century psychology experimental research practices and to describe the way research was replicated.

The original experiments I present in this paper are well known, dating from 1860 to 1900, a period characterized by important changes in Europe, such as industrialization, workers movements, and the constitution of modern nation states. In this period, the Prussian model at German universities (Charle, 2004), offered a broad humanistic as well as a thorough experimental training in fields such as chemistry, physics, physiology, philosophy, and psychology, which was just emerging as an experimental science with its own scientific community (Ash, 1980; Bringmann & Tweney, 1980; Rieber & Robinson, 2001). When the first replications were performed, neither young researchers nor senior professors suffered the pressure to “publish or perish,” and journals were not yet acting as gatekeepers taking decisions based on *p*-values. Often, researchers carried out years of patient experimenting before they published their results. Thus, what role did replication play within psychological research?

## What Is Scientific Replication and When Did it Start?

Replication is a rather recent term that has been used by scientists since 1914 to refer to the activity of repeating an experiment or a series of observations (Tweney, 2018).<sup>3</sup> In

<sup>1</sup>Theory & History of Psychology, University of Groningen Faculty of Behavioural and Social Sciences, Groningen, Netherlands

### Corresponding Author:

Annette Mülberger, Theory & History of Psychology, University of Groningen Faculty of Behavioural and Social Sciences, Grote Kruisstraat 2/1, Groningen 9712 TS, Netherlands.  
Email: [a.c.mulberger@rug.nl](mailto:a.c.mulberger@rug.nl)

several ways, the practice is connected to broader issues in the history and philosophy of science. The replication of experiments is deemed important by scientists as a way to ensuring stable and generalizable results. Scientists (i.e., “natural philosophers”) have been aware of the need to repeat experiments for centuries. Schickore (2011) showed how redoing another investigator’s experiments became an issue around 1670 and Grower (1997) pointed to the role this played in early scientific debates concerning Newton’s experiments.

Nevertheless, since the 1980s, experimentation and replication have been criticized by sociologists, philosophers, and historians of science and interpreted as being the result of a rather naïve commitment to objectivity as an epistemic ideal (e.g., Daston & Galison, 2007). Shapin & Schaffer (1985) considered replication to be the product of the set of technologies which transforms what counts as belief into what counts as knowledge; it includes physical reiteration of a certain kind of experimentation as well as virtual witnessing through “literary technology.” In the following, I will expose some examples of tables and graphs used by psychologists.

Although physical or virtual “reiteration,” “repetition,” or “replication” of an experiment sounds like an easy task to perform, judging whether an experiment is a valid copy is highly problematic because there is no unambiguous set of rules. Moreover, in his book “Changing order,” Collins (1985) pointed out the problem of experimenter’s regress. He views a scientific fact as an expression of a “form of life” (in Wittgenstein’s sense). In a controversy, each group argues from a different stance or “world view”; there is no algorithm that can be used to evaluate different views objectively. “Experimenter’s regress” refers to a closed loop: researchers need to accept, as a matter of fact, the phenomena of the experiment they want to replicate because they need them to calibrate their own machines or methods; only when they obtain the same result can they be sure that their machines or methods have worked.

If opinions as to the success or failure of the replication vary, Collins and other sociologists of science argue that there is no way to decide who is right. The epistemic uncertainty in experimentation, involving unexpressed and partially reported skills and assumptions, makes it impossible to fix clear-cut criteria to decide whether an experiment constitutes a “genuine” replication or not. Nevertheless, Radder (1992) argues that in most cases, the regress can be overcome with the help of negotiations about the adequacy of the instrument and the kind of phenomena produced by them.

Up until this point, I have cited works referring to replication in scientific practice, in general. What about this practice in psychology? Did nineteenth-century psychologists actually consider it important to replicate experiments? The answer is not self-evident, as publications on the replication crisis in psychology focus only on recent decades (see, for example, Wiggins & Chrisopherson, 2019). Laws (2016) even talked about an “enduring historical abandonment of replication” (p. 2). In order to clarify just when psychologists started to examine replication, Makel, Plucker,

and Hegarthy (2012) performed a bibliographic search that showed increased use of the term “replication” in scientific publications since the 1950s. Nevertheless, this cannot be taken as an indicator of the real number of replication studies because previously alternative expressions were used instead of “replication” (for example: “repetition,” “reproduction,” “redoing,” and “testing”).

Precisely dating the historical origin of replication in psychology is not easy. Pettit (2018) has developed a timeline which chronologically situates a variety of milestone experiments and their successful or unsuccessful replications, starting with the examination of Anton Mesmer’s magnetic (hypnotic) therapy at the French Academy of Sciences in 1784. Danziger and Shermer (1994) offer the only historical analysis dealing with practices of replication among psychologists of the past. They found that in the late 19th and early-twentieth century, investigators “(...) regarded each additional subject exposed to constant experimental procedures [in their own research,] as a replication of the original experiment” (Danziger & Shermer, 1994, p. 21). Psychologists generally disagreed on replications and the replicability of each other’s experiments. Danziger and Shermer (1994) demonstrate this via the Wundt–Bühler controversy (1907/08) concerning the value of experiments on thinking and the Baldwin–Titchener controversy (1895/96) over the criteria for the selection of experimental subjects. The two historians conclude that the issue of what constitutes a valid object of study (in the first case) and a valid experimental subject (in the second) lay at the heart of these controversies.

Following Collins and Hacking, Danziger and Shermer (1994) attributed the clashes over replication to divergent styles of scientific practice. Wilhelm Wundt (1832–1920) and Edward Bradford Titchener (1867–1927) drew their models for proper research from the natural sciences: physics and physiology. For them, experimentation involved the creation of artificial laboratory conditions offering a high degree of “purity,” which could only be achieved by carefully selecting experimental subjects and isolating certain features or variables. The price of this “purity” was severe restrictions on the objects and scope of research. Bühler and Baldwin, in contrast, adopted a looser and more openminded approach. Bühler accepted a mental state triggered by an intellectual task or puzzle as a valid object of study. Baldwin argued for using a sample of experimental subjects taken “from the given, everyday world,” instead of following Wundt’s rule of carefully selecting trusted insiders (who might be biased).

Nevertheless, there are several problems relating to the historical analysis undertaken by Danziger and Shermer (1994), as I will show in the present paper. To begin with, their study does not warrant at all their general concluding claims about psychologists assuming that their research methods are independent of the object of study. Moreover, the fact that human beings are historical entities, changing over time, does not, in principle, preclude investigation of some relatively stable, shared, or regular psychological features or

processes that can be identified and studied via (replicated) experimentation.<sup>4</sup>

My aim is to examine in more detail the way replication worked in early psychological experimentation. On a methodological level, my approach differs from previous research in several ways. I start my narrative earlier, focusing on some classical psychological experiments from the period 1860–1900, a key period in the emergence of psychology as an experimental science. Moreover, and contrary to Collins' claim that in science replication rarely worked and to Danziger and Shermer's work on controversies and failed replications, my historical examples include some successful repetitions. Nevertheless, in line with Collins (1985) and Shapin and Schaffer's (1985), my examples also show that agreement on the delimitation criteria between a "genuine" or "correct" replication and a "mistaken copy" becomes problematic as soon as we move into a controversial setting. This is even the case when both researchers share their model for proper research and many salient features of their laboratory routine. Furthermore, the present analysis of historical case studies evidences a rich variety of functions. I argue that replication constituted a tool not an end; it was a means of acquiring expertise and group identity, providing knowledge of and acquaintance with specific individual practical features, as well as a way to discuss and even attack theories held by well-known authorities.

## Psychophysics

### *Fechner's Experimental Praxis*

Psychophysics was developed by Gustav Theodor Fechner (1801–1887), a physicist who worked on electricity and optics. In the 19th century, in fields such as optics, a new kind of experimentation became common, centered increasingly on quantification and measurement (Buchwald, 1989). Hacking (2014) showed that this was a general trend that reached the social sciences and psychology, stating: "the world was becoming numerical." Statistics were increasingly used to record and treat census figures (Porter, 1995), and toward the end of the century became a way to model variation and to infer general trends from numerical data (Gigerenzer et al., 1989). Fechner was aware of the work of Laplace, Bernoulli, and Gauss and contributed substantially to the development of statistical methods. Keen on quantification, he used the error law in his optics work and later extended its use to his psychophysical experiments. Furthermore, in the debate about determinism, he adopted an indeterminist stance, which he later tried to support mathematically (Hacking, 2014; Heidelberger, 2004).<sup>5</sup>

Within his broader philosophical (panpsychist) project<sup>6</sup> and inspired by the work of the physiologist Ernst Weber, Fechner defined psychophysics as an "exact doctrine [dealing with] the functional, or interdependent, relation between body and soul (...)"<sup>7</sup> (Fechner, 1860, p. 8). In order to explore

this relation, he experimented on himself using three main methods<sup>8</sup> of which the method of "adjustment" was one, determining the just noticeable difference (JND). Thus, a subject is asked to select the level of intensity of a stimulus (a weight, sound, or light) that it is just barely detectable or at the same level as another stimulus. Fechner took as his starting point the "principle of insufficient reason"<sup>9</sup> which assumes that random variations can be compensated through numerous repetitions of experiments. Thus, he stated that irregular "casualties [must be] compensated through frequent repetitions in a way that if the variation and sensibility stay the same, one obtains coincident results in the measurements taken at different times; this way the individual casualty loses weight and the final results are in so far independent from chance" (Fechner, 1860, p. 79; the emphasis is Fechner's).

To gather the high number of empirical observations needed for his calculations, Fechner experimented regularly at his home in Leipzig. "For several years [since 1855]," he explains, "I considered it as part of my daily work to undertake experiments during 1 hour (...)" (Fechner, 1860, p. 93). Psychophysical methods require working with formulas and measurement and much time and patience collecting data. For example, in the years 1856 and 1857, he explored his perception of a series of weights using the method of adjustment. He needed 24,576 experiments in which he recorded his psychological appreciations of weights with the "real" weight of the stimuli (measured in grams). On the whole, his measurements led him to conclude that the magnitude of sensation corresponds to the logarithm of the magnitude of the physical stimulus.<sup>10</sup> In the following 2 years (1858–1859), he completed another 16,384 experiments. After comparing both series, he concluded satisfied: "(...) the main result of this [later] series constitutes a complete confirmation of the previous results" (Fechner, 1860, p. 196).

### *Criticisms and Appropriations*

Although Fechner did most of the experiments by himself at home, he did not work in isolation. Because of his bad eyesight, he needed others to do the experiments with lights. Thus, Fechner (1860) explained how some of the experiments were performed and repeated by his brother-in-law, the physiologist Alfred Volkman, as well as by other colleagues (physicists and physiologists) who obtained similar results.<sup>11</sup>

Following Heidelberger (2004), "Fechner's formulas and methods had unleashed enough 'paradigmatic energy' to start-off a new, normal, scientific tradition" (p. 212). Whether we can speak here of a specific paradigm or not is questionable, but it was clearly a challenging project which in the 1870s triggered strong reactions from philosophers, physiologists, and psychologists. Fechner's idealist approach clashed with the then growing materialism (Heidelberger, 2004). The philosophers Bergson and Mach questioned Fechner's attempts to measure sensations, arguing that the

intensity of a sensation is not gradated, composed by a sum of psychological (JND) unities.

Psychologists such as Wundt and Titchener, together with Hermann Ebbinghaus, Marcel Foucault, and Georg Elias Müller also criticized Fechner's work<sup>12</sup> and rejected his spiritualist stance. At the same time, however, they by replicating his experiments adopted and tried to improve his methods. They viewed psychophysics as an innovative method for psychology. It required the use of some physiological instruments, as well as mathematical (statistical) skills; mastery that was not acquired via philosophical courses. For psychologists, it became a way to obtain experimental knowhow, as well as of enabling them to distinguish themselves professionally from (non-psychologist) philosophers and align their research methods with those of natural scientists.

Psychophysics became disconnected from Fechner's philosophical project and turned into an empirical science, attempting to reveal replicable trends in human perception.<sup>13</sup> Viewed as a scientific (objective) method, psychophysics was attractive to scholars from very different cultural, political and religious backgrounds who replicated Fechner's experiments. Two examples follow. Apart from the well-known psychologists cited above, mostly with protestant backgrounds, also the German Catholic priest Constantin Gutberlet (1837–1928)<sup>14</sup> became interested. Eager to understand the working of the new “science of the soul,” he started to replicate psychophysical experiments. His aim was to discuss its findings and judge the worth of its theories. This was rare among Catholic priests; the conflict (*Kulturkampf*) of the 1860s, when Bismarck's policy clashed with the Vatican's interests, had placed them in a difficult position. After 1870, the situation improved with the papal publication of *Aeterni Patris* in 1879, encouraging Catholic thinkers to embrace modern science. This was taken up by the Görres-Gesellschaft, whose influential journal (“Philosophische Jahrbuch der Görres-Gesellschaft”) was edited by Gutberlet between 1888 and 1924. Despite the fact that neo-scholasticism became an international movement, it was still controversial for a priest to be actively involved in an experimental science such as psychology. Gutberlet (1905) prepared a critical and comprehensive exposition of psychology's main findings and principles. On the one hand, he presented the empirical results together with current psychology theories and debates. On the other, he added corollaries to the theoretical discussions, arguing from a neo-scholastic point of view against the prevalent “materialism” in psychology as well as Fechner's panpsychism (see Gutberlet, 1905, p. 191–193).

Another promoter of psychophysics was the younger psychologist Martín Navarro Flores (1871–1950), who worked in Catholic Spain. He was a member of the pedagogical institution “*Institución Libre de Enseñanza*” which endorsed an idealist-positivist philosophy (called Krausism) and launched a progressive, freethinking, educational reform movement.<sup>15</sup> In his textbook on experimental psychology published 1915, he dealt in depth with sensation and

perception, frequently citing Fechner's work (Carpintero, 2004). After lamenting Spain's backwardness, Navarro Flores (1915) invited his colleagues to engage in psychological experimentation by replicating psychophysical experiments (Lafuente, 1988). The assumption about cultural and national differences influencing the outcome of psychological experiments and mental testing was widespread at that time in Spain (see, for example). We need to explore how the Spanish people will react to the tasks; he argued (Navarro Flores, 1915). Thus, instead of reproducing Fechner's results, Navarro expected psychophysical experiments to evidence the idiosyncratic way “the Spanish mind” perceives the world. He did this despite criticizing Fechner for not having considered further the role of physiological (bodily) processes as a bridge between physical and psychological phenomena (Lafuente, 1988). The strong neurophysiological tradition of Cajal's school in Spain at that time placed much emphasis on the brain functions, triggering his interest towards what Fechner called “inner psychophysics.”

In short, Fechner's panpsychic theory was mostly rejected by psychologists. His psychophysical experiments were also polemical and bombarded with criticism. At the same time, psychologists such as Wundt, Titchener, and Ebbinghaus welcomed psychophysics as a new way to experiment with psychological processes. They replicated Fechner's experiments, motivated by the hope to find a stable trend in the repeated measurements that would lead to knowledge about the functional relation between the outer world and human perception (i.e., the inner world). While the praxis of psychophysical experimentation acquired significance signaling the identity of an emerging scientific community, later called “experimental psychology,”<sup>16</sup> replication of Fechner's methods was also undertaken and promoted for different purposes outside the inner circle of German experimentalists.

## Memory Experiments

### *Ebbinghaus' Memorizing Routine*

The innovative memory experiments conducted with great care in the 1880s in Berlin by Hermann Ebbinghaus (1850–1909) were celebrated by Georg Elias Müller, William James, and Edward B. Titchener as a great advancement because for the first time a “central [higher] psychological function” had been experimentally investigated<sup>17</sup> (Müller & Schumann, 1894; Shakow, 1930). Ebbinghaus promoted psychology as a natural science and was, therefore, considered to be a “materialist.”<sup>18</sup> He was a rather independent researcher, strongly influenced by Fechner's psychophysics.<sup>19</sup>

Ebbinghaus' aim was to study the effect of time on rudimentary memorizing. In 1883, he started exploring the time conditions under which he could learn and reproduce series of syllables without error (Ebbinghaus, 1885).<sup>20</sup> His contribution is widely known,<sup>21</sup> thus, I will summarize it only very



**Table 1.** Table summarizing one of Ebbinghaus' Main Results: The Amount of Time Saved When Relearning Syllables (Source: Ebbinghaus, 1885, p. 103).

Nr.	I After X hours	II [the experimental subject] was able to remember this much, so that when relearning the lists a saving of Q% of the original learning time was achieved.	III Wm (average value/ Mittelwerte)	IV The amount that was forgotten corresponded to an effort (Arbeitsleistung) of v% of the original [effort when learning the syllables for the first time].
	X =	Q =	Wm =	v =
1	0,33	58,2	1	41,8
2	1	44,2	1	55,8
3	8,8	35,8	1	64,2
4	24	33,7	1,2	66,3
5	48	27,8	1,4	72,2
6	6×24	25,4	1,3	74,6
7	31×24	21,1	0,8	78,9

briefly: Using himself as the experimental subject, he read aloud lists containing varied numbers (12, 24, 36, etc.) of “nonsense syllables” at a rapid rate of 150 syllables per minute. When he finally had the impression that he had managed to memorize them, he tried to repeat the series by heart at the same pace. Whenever he noticed a mistake or a hesitation, he resumed reading until the learning was successful.

Ebbinghaus evaluated the effort (“Arbeit”) needed for the learning indirectly, by measuring the number of readings and the time required to learn each list (Ebbinghaus, 1885, p. 41). To do this, he did not use any sophisticated instruments: only the lists of syllables, a metronome, a watch (or a chronoscope), and a rope with wooden balls for keeping count. One of his principal results was: learning a longer list requires more repetitions and relearning a list “costs” less time than learning it for the first time; the more time passes, the less is the advantage (see Table 1).

The second column shows the time that has passed since the first learning of the syllables list. The percentages in the third column show that relearning after 20 min was the most efficient, requiring less time (in comparison to the time needed to learn the list of syllables for the first time). This advantage decreases as the time interval between the moment of learning and relearning increases. After 1 hour, for example, the percentage of time saved decreases from 58.2% to 44.2% and steadily decreases until reaching 21.1% after 31 days (see last percentage in the third column). The last column shows the complementary percentage of the forgetting. Similarly to Fechner, Ebbinghaus argued that these measurements conform to a logarithmic formula which can be represented as curve.

Ebbinghaus seems to hold economic interest, referring in his report to “mental work” and “time savings,” concepts that reflect the industrializing period in which work (force), economic efficiency, and soon also scientific management (Taylorism) would impose minute regulations to increase productivity rates. Danziger (1987) interpreted his terminology as reflecting an “energetic model,” in which the level

of psychological performance is viewed as a consequence of the mental energy accumulated and the amount of work invested in the memory task.

Even more striking is the fact that Ebbinghaus' report also examined in detail limitations and possible sources of error. He justified the fact that his experiments did not reflect the complexity of everyday life by reference to physicists who also work with abstractions. He was concerned, nevertheless, about the way in which changes in real life distorted the outcome of his measurements, such as those resulting from uneven material, changes in attention or mood, etc. Thus, Ebbinghaus adopted two strategies to ensure the objectivity of his data. Just as with Fechner, he expected the functional relation (*Abhängigkeitsverhältnis*) to be relatively constant whenever a large number of nearly mechanical repetitions were performed, in which oscillations would be compensated. Thus, he repeated his experiments over and over again, imposing a tough experimental routine on himself and a regular lifestyle.<sup>22</sup>

When he undertook a self-examination of his own mind, as a researcher, he noticed a dangerous source of error: his own expectations, or, as he said it: the “secret influence of theories and points of view” (Ebbinghaus, 1885, p. 41). Even if he consciously tried to counteract such a trend, this would alter the natural working of the mind and thereby the outcome of the experiments. As a remedy, he made an effort to ignore the results as much as possible, while doing the experiments, and to examine them carefully and critically once experimentation was over. Such reflections might explain the rather curious strategy he adopted to add further proof to his results, citing previous “control-experiments” (from 1883/84, see Ebbinghaus, 1885, p. 107). He probably felt that these offered better empirical support because he was not yet aware of the results of his main study. Furthermore, he suggested using another person (who is unaware of the aim of the study) as the experimental subject in future research.

Ebbinghaus' (1885) examination of consciousness not only identified a dangerous source of bias, but also detected a

**Table 2.** Example of Müller and Schumann's (conceptual) Replications of Ebbinghaus' Experiments: The Influence of the Rhythm of Intonation on the Learning Time (Source: Müller & Schumann, 1894, p. 157).

	$W_a$ :	$W_c$ :	
trochaic learning	19.0	18.5	(n = 48)
jambic learning	20.3	20.0	(n = 48)
trochaic re-learning after trochaic learning	8.7	7.8	(n = 24)
jambic re-learning after trochaic learning	10.5	10.0	(n = 24)
trochaic re-learning after jambic learning	10.4	9.0	(n = 24)
jambic re-learning after jambic learning	8.9	7.4	(n = 24)

" $W_a$ " refers to the arithmetic average  
" $W_c$ " to the median ("Centralwert").

will to learn about the real outcome of the experiment. He stipulated that, in the long run, these contrary trends would probably become compensated: whenever performance might become distorted by some involuntary desire to enhance a certain effect, it is probably balanced out by other trials in which the opposing desire to discover "factual truths" prevailed. One would not want to put so much effort into one's work, he reasoned, just to base it on the weak finding of "one's own phantasy" (Ebbinghaus, 1885, p. 41).

Ebbinghaus' self-analysis is interesting. We can see that, despite his commitment to positivist epistemology, he was aware that desires and expectations might condition the working of his mind. Ebbinghaus optimistically hoped that with the help of some cognitive strategies and numerous replications through which he aimed to rule out the two opposing biases (desires). Thus, he applied exactly the same statistical reasoning to the working of his mind as he deemed a scientist would, in perfect analogy to the ruling out of confounding variables in his experiments.

### Early Reactions and Müller and Schumann's Replication

Ebbinghaus' (1885) work received praise as well as criticism in Germany and abroad. William James was impressed and a book review in the British journal *Mind* underlined his rigor and patience (Jacobs, 1885).<sup>23</sup> One of the most detailed examinations and repetitions of Ebbinghaus' research on memory was performed by Müller and Schumann (1894).<sup>24</sup> Their aim was "to get acquainted with Ebbinghaus' methods" and "to contribute to their improvement, accurateness and extension (...)" (Müller & Schumann, 1894, p. 81). At the same time, they hoped to gather new insight into the workings of memory. Over the 4 years from 1888 to 1892, Müller and Schumann organized 13 series of experiments (each running for approximately 100 days) in which they mechanically controlled the time of exposure of each syllable with a rotation apparatus (Rotationsapparat) and adopted several strategies to make the series of syllables more homogeneous. Furthermore, in some experiments, they systematically varied a variable. For example, in one series, they changed the rhythm of intonation to study the effect of this on the learning process. Thus, the experimental subjects used either

a trochaic or an iambic rhythm, a variation which did indeed alter the results, as can be seen in Table 2.

The list of averages shows that the time required for the learning is similar using either rhythm, being for the trochaic type slightly lower (see first two lines in Table 2). Moreover, when the syllables were learned using one rhythm and afterward had to be re-learned with another, the time savings were less (see lines 4 (jamb-troch) and 5 (troch-jamb) than in the cases in which the rhythm had been maintained (see lines 3 (troch-troch) and 6 (jamb-jamb).

As in the original experiments, also for Müller and Schumann's replications much patience was needed to complete the many hours of experimentation. Aware of the confounding variables that could distort objective results, such as changes in mood, bias due to expectations, and the like, they followed Fechner's idea of balancing out errors through replicating a large number of trials. Müller was known as a drillmaster, imposing a tight regime of rules and checks on his students working at his laboratory. Kusch (1999) described his memory experiments as a parade-ground drill in which "(...) the actions of the subjects were highly constrained, repetitious, and somewhat mindless. Usually, the subject learned nonsense syllables without knowing the purpose of the experiment. The parallel between the parade ground and the memory experiment is strengthened further by the vocabulary used in the context of the latter: 'full hits', 'partial hits', 'drum', and 'sacrifice'" (p. 106). Here, it might be relevant to point out that Ebbinghaus and Müller had military experience in the French-Prussian war (1870/71).

Müller and Schumann's major innovation, though, was to separate the role of the experimenter from that of the experimental subject. They carefully checked subject reliability, selecting only those "for whom we could presuppose full reliability and a love for the truth" (Müller & Schumann, 1894, p. 264). But finding experimental subjects was not easy. They explained: "(...) memory experiments, as we have undertaken them, demand a great amount of patience and sacrifice of time and freedom in lifestyle, which is not agreeable to everyone" (Müller & Schumann, 1894, p. 264, p. 264).<sup>25</sup> Thus, control was not only executed within the experimental setting: similarly, to the work of Ebbinghaus, Müller tried to control external conditions as much as possible, asking the subjects to maintain a constant lifestyle.

Despite ensuring uniformity through rigorous repetitions of the procedures, Müller and Schumann were also aware of the existence of individual differences. Influenced by the work of Binet, they recognized the specific “sensorial character” of one person’s memory, referring to the fact that some people seemed to remember images better, while another reacted more to sound. They argued that specifically the repetition of a standardized task over a long period of time brings the idiosyncrasies of the subject’s mind to light. Nevertheless, after a quick comparison, Müller and Schumann (1894) arrived at the conclusion that individual differences are less relevant than other variables.<sup>26</sup>

### *Replications a Century Later*

Despite Müller and Schumann’s successful replication of Ebbinghaus’ experiments, in the past and in the present, critical voices have questioned the value and the validity of Ebbinghaus’ legacy. Reductionistic mechanization and mathematization for experimental purposes is considered highly problematic when it comes to such a complex and meaningful process as memory. Ebbinghaus and psychologists working on memory after him were conscious of the limitations of his insights referring only to a very rudimentary, highly trained kind of memorizing process.<sup>27</sup> Moreover, mechanical learning soon became demonized by pedagogues and psychologists; thus, Bartlett (1932) famously referred to Ebbinghaus’ learning experiment as being based on irrelevant “repetition habit.”

Notwithstanding such criticisms, Ebbinghaus’ memory research had a long-lasting impact on psychology, appearing regularly in psychological textbooks. In 1985, when the centennial of his publication was celebrated in Passau (Traxel, 1987), historians such as Danziger described “Ebbinghaus’ pioneering work” as a “fundamental contribution to the development of modern psychology” (Danziger, 1987, p. 217). He viewed Ebbinghaus’ success in the applicability of his energetic approach to fit in well with educational concerns. Other authors such as Van Rappard (1987) stated that, although memory research had changed over time, “it can safely be said that he set the tone for what may be called in terms of Larry Laudan (1977) a ‘research tradition’” (p. 43–44).

But not all historians would value Ebbinghaus’ contribution. Draaisma (1995), for example, instead of considering it the beginning of a successful experimental tradition in psychology, laments the experiments as marking the end of an era of Romantic literary and neurological traditions. At the negative extreme of the spectrum of opinions, we find Smedslund, who pictured Ebbinghaus as a clever “illusionist,” disqualifying his experimentation with the trivial accusation of not having kept track of the infinite number of different sources that might potentially have influenced the findings (Smedslund, 1987).<sup>28</sup>

Despite its shortcomings, psychologists, on the other hand, often took Ebbinghaus’ forgetting curve as reference point. Tulving (1985) claimed that replicating Ebbinghaus’ work had become increasingly difficult over time because “most other

people, especially in today’s world, are probably incapable of mastering any longer lists of nonsense syllables under [Ebbinghaus’ experimental] conditions” (p. 486). Whether unwilling or incapable, patient experimental subjects prepared to submit themselves to such a mechanical learning drill were never easy to find and even when they could be found, replication was cumbersome and problematic.<sup>29</sup> Nevertheless, a century later, some researchers managed to successfully replicate Ebbinghaus main experiments, following his instructions very closely (Heller, Mack, & Seitz, 1991; Murre & Dros, 2015). The aim of the replications was to “verify the reliability” of the original results and to “uncover” how the experiment was conducted (Murre & Dros, 2015, p. 2).

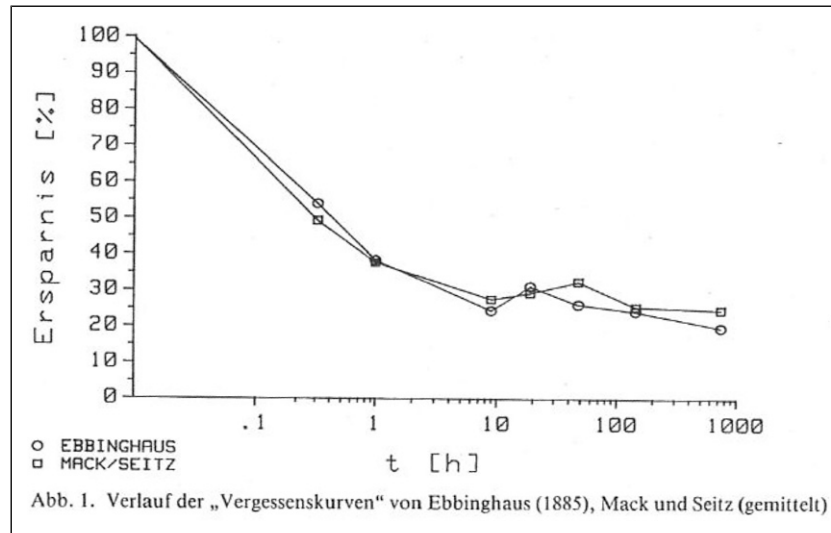
The graph in Figure 1 below shows Ebbinghaus’ results as well as the averages obtained in the replication with regard to the percentage of time savings in relearning process after different time intervals. Despite slight differences between the two curves, the same trend can be detected: the more time has passed since the moment a list of syllables had been learned, the more time is needed to re-learn it. Thus, when using such strategy to visualize replication, each attempt is represented as additional layer to the original graph (see also the three layer graph in Murre & Dros, 2015).

While other previous attempts did not manage to obtain similar data and had struggled with a problem of interference, Heller, Mack, and Seitz (1990) noticed that Ebbinghaus’ original learning speed—reading 150 syllables per minute—efficiently inhibits any mental search for associations or interference between them. Thus, to be able to read the material at the prescribed pace, thorough preparation was required. Only after hours of training is it possible to obtain a regular level of learning and relearning. Similarly, Shebilske and Ebenholtz (1971) had observed that problems in replicating the original results could be attributed to the fact that “Ebbinghaus was a highly trained learner whereas most modern experiments have used naïve subject” (p. 555). Given the pre-trial training and the way Ebbinghaus set up the experiments, he was certainly not studying memory as it is generally understood or used in everyday life. Despite the shortcomings in scope and external validity, it seems clear that his work offered some new and interesting insights into the working of the mind, inspiring many psychologists after him.

## **Controversial Emotions**

### *Replication and Self-Perception of Emotions in Wundt’s Laboratory*

To facilitate the setting up of psychological experiments such as Fechner’s and Ebbinghaus’, in 1879, a psychological laboratory was established by Wilhelm Wundt in Leipzig and, soon after, other psychological laboratories were founded all over the world. These were places where students could specialize in psychology. Thus, in the last two decades of the 19th century, a generation of psychologists received systematic training, giving rise to a new community of experts



**Figure 1.** Learning and Forgetting After a Century: A Comparison of Two Curves of Forgetting, the First Obtained by Ebbinghaus in 1885 and the Second by Mack and Seitz in 1991. (Source: Heller, Mack, & Seitz, 1991, p. 9; reproduced with the permission from the current publisher: Hogrefe Publishing.).

(Ash & Geuter, 1985; Danziger, 1990). In this setting, we find another use of replication: experimental psychology became a collaborative enterprise, and the repetition of experiments was part of student training and of learning laboratory routines in hierarchically organized laboratories in which teaching and research went hand in hand.<sup>30</sup>

Experimental psychology constituted a productive, collective, and increasingly technical undertaking.<sup>31</sup> Research became more limited in time, as students usually finished their PhD in one year. Although not all PhD research was experimental, to replicate psychophysical experiments became a way to assert one's professional identity, demonstrating knowhow that distinguished a psychologist from a (traditional) philosopher. In Leipzig several methods were employed, such as psychophysics, mental chronometry, and introspection.<sup>32</sup> Introspection was a problematic method, though in the form of self-perception, it was viewed as a necessary tool because it provided access to psychology's object of study: the human consciousness. Thus, Wundt declared: experimentation “enables us to repeat the subjective sensations and emotions, which come along with the process as often as we wish (...)” (Wundt, 1888, p. 433).<sup>33</sup>

While psychologists started to have extensive experience measuring sensations, identifying, and measuring emotions<sup>34</sup> was a more difficult task. Also, the cognitive status of emotions and their relation to bodily functions was a controversial topic. Wundt's physiological measurements of emotions, together with the James–Lange theory (1884), stipulated the relevance or even priority of physiological changes in emotional reactions. Following Dror's (1998) research, the appeal of physiological approaches would reach its peak soon after, in 1906.

Throughout his life, Wundt changed his theory several times (Wassman, 2009). In the *Grundriss* (1896), he

developed his “voluntarist psychology” and presented emotional experience as being mixed, situated between three bipolar axes: pleasure–displeasure (Lust–Unlust); excitement–inhibition (Erregung–Beruhigung); and tension–relaxation (Spannung–Lösung).<sup>35</sup> He combined introspective experiments with registering the concomitant physiological variables,<sup>36</sup> which constituted, in his view, “the regular symptoms of the emotional processes” (Wundt, 1896, p. 102). Thus, for example, a weak pulse that accelerates would usually accompany displeasure, while a strong pulse that slows down would indicate pleasure (see Wundt, 1896, p. 103).

### Titchener's Critique

Replication becomes even more difficult when experiments were repeated in order to criticize or disqualify some theory, method or findings, leading to a scientific controversy. In the introduction, I mentioned the study by Danziger & Shermer (1994) on replications. They indicated controversies between two groups of scholars: on the one hand, Wundt and Titchener, grouped together as “purists” who aimed to emulate the natural sciences and whose way of practicing experimentation clashed with that of Bühler and Baldwin who, on the other hand, represented a looser conception of psychology. I will show in the following section that this description is problematic because controversies occurred even among the “purists”, that is, researchers working within the Wundtian line of experimental psychology. Again, replication constituted a core element within the debate.

In 1899, Titchener, Wundt's most faithful follower, published a critique of Wundt's theory on feeling. He offered “empirical facts” (Tatsachenmaterial) contradicting the master's three



dimensional theory, which had already become highly contested among psychologists. Titchener's new facts had been gathered by one of his students ("Herr W.").<sup>37</sup> The result of his replication was negative: the student could only find his feelings varying from pleasure to displeasure, neither of the other two dimensions of Wundt's theory appeared in his introspective reports.

The failed replication by Titchener's student constituted a public disqualification of Wundt's research. As one contemporary observed: "Thus the patient thinking of the expert Wundt is brought to zero with the greatest dispatch (...)" (Buchner, 1900, p. 96). Wundt (1900) was offended and immediately questioned the validity of Titchener's critique and his student's replication. He denounced several methodological errors as reasons why the results differed. First, he criticized the fact that Titchener had left it up to a student to undertake the self-perceptions. Nevertheless, this was not a mistake: it indicates differences in research practices between the two laboratories. Whereas in Leipzig, professors and expert psychologists (for example, Wundt himself or one of his assistants) acted as experimental subjects, at the Cornell laboratory, trained student were used (not the author whose theory was being tested).<sup>38</sup>

Second, Wundt criticized the omission of the records of the underlying bodily reactions. This again was not neglect: such physiological measurements were deemed useful in Leipzig where Wundt claimed that bodily symptoms were the parallel expressions of feelings; but in Cornell, this theory had been rejected two years earlier by another of Titchener's students, David Irons (1897). Irons had shown that bodily responses were only incidental to emotions, not intrinsic features. Thus, for Titchener and his students, such registers were not at all informative.

In 1908, Titchener, confirmed his critical stance once more. Visibly disturbed by the confrontation with his former teacher, Titchener ended with the following ethical rules, phrased in a prayer-like rhetoric: "(...) we must not be dogmatic, we must not be [sic!] too impatient for results, we must not set theory above observed facts: (...) we must use all the weapons in our critical armory against ourselves as against others, and against others as against ourselves" (Titchener, 1908, p. 231). This was, again, a resounding criticism of Wundt's research method and dogmatic attitude. The main "weapon" of that "critical armory" Titchener had in mind was precisely replication in the form of empirical results that would speak for themselves.

Toward the end of the 19th century, Wundt found himself in a scenario including emancipated former students. Tensions between competing researchers and laboratories polarized the variety of psychologies practiced at the time. Thus, this episode can be viewed as a single chapter in a more extensive history of repudiations against Wundt (Danziger, 1979; Mülberger, 2012). At the same time it also shows that Danziger and Shermer's grouping together of Wundt and Titchener as "purists" is problematic. It demonstrates that their categories of analysis are not adequate when it comes to

explaining the underlying split leading to controversies and failed replications. If, as they state on page 22, both Wundt and Titchener's model of proper research was derived from the experimental sciences (namely physics and physiology), then why should they clash over a failed replication of each other's experiments?

The replication of Wundt's research on emotions by Titchener's student evidences disagreement that existed within the Wundtian tradition, involving two levels: theory and research practices. While Titchener adopted many salient features of Wundt's laboratory practices (Boring, 1927; 1950), there were also some striking differences between their philosophies of science, theories, and working styles. Titchener rejected Wundt's (parallelist) body–mind theory, as we have seen, and his epistemological stance was closer to that of the British empiricists as well as Mach (Araujo & Marcellos, 2017; Leahey, 1981).

Thus, in the historical context of competing psychological laboratories, replication acquired yet another social function: challenging claims made by established authorities. The empiricist, Titchener, was one of Wundt's most faithful students and was certainly not keen on having a personal confrontation with his former mentor. His strategy was to let the "objective" (empirical) results of this student's research "speak for themselves." Nevertheless, Wundt was well known for reacting with fiercely personal attacks whenever his work was criticized: here he would make no exception.

Titchener's strategy was not uncommon. In the 1890s, Mary Calkins and her PhD student Cornelia Nevers challenged Jastrow's study of word associations, in which he had examined group differences (Mülberger, 2017). Jastrow had recognized distinctive trends in the responses when comparing men's reactions to women's, attributing some differences to women's "household instincts." With the help of several repetitions, the Wellesley psychologists offered different data, insisting that even if there were minor differences, these could not be attributed to any biological differences between the two groups (see also García Dauber, 2005).

## Conclusions

In the present paper dealing with the praxis of replication, this is broadly understood as repetition of one's own or others' experiments. I have reviewed replications in three types of classical experiments that were undertaken during the last decades of the 19th century. They do not constitute all kinds of psychological research that were performed at that time, but they do represent a kind of research in which replication played a prominent role.

Overall, the early experimentalists' psychological research can be viewed as a kind of small scale and slow science, if compared with current practice, in the sense that it implied a reduced number of experimental subjects and years of repetitive experimentation. My analysis suffices to show that

replication was an issue in the work of these experimentalists and, thus, has a longer tradition in psychology than current publications seem to imply.

The core idea when seeking confirmation was that psychological measurements under experimental conditions could indicate some stable, functional relations between two variables. In order to rule out other (undesired) influences, Fechner and Ebbinghaus repeated their experiments thousands of times and invited other researchers to do further replications. Thus, the first role of replication was to balance out error: to counteract the disturbing effects of confounding variables. Moreover, the findings, patiently acquired with the help of a few experimental subjects, were expected to help understand the workings of human mind. Any additional subject going through the same experimental conditions represented a replication.

In the psychological publications of the 19th century, findings (including formulas and data) were generally explained in the text. Additionally, some main results (usually averages for each experimental subject or experiment) were listed and presented in tables. Graphs were more rarely employed, posing difficulties regarding the representation of exact amounts. Replications were often represented within tables and graphs as an additional layer, adding a column or curve for each attempt. Such a “literary device” facilitated comparison.

No historian doubts that human beings and society at large have changed since the 19th century. At first sight, such an awareness seems to deny any useful role for replications across time and history. Moreover, there is no standard human mind but only very different and unique people. Given these assumptions, how can we expect psychological findings ever to be stable? Historians and philosophers of psychology nowadays generally adopt a negative position and often reject the value of experimentation and replication altogether. But in the light of the present research, this seems to be an overstatement, as there are psychological findings that have been successfully replicated. An experiment such as Ebbinghaus’ can be reproduced yielding similar results, even after a century. This seems to indicate either a certain level of generalization within the workings of certain psychological processes across individuals and time, or a certain stability of social conventions. In any case, the fact that humans and their worlds change over time does not *per se* preclude historical-psychological continuity and the possibility of replication fulfilling different roles.

The historical cases of my research demonstrate further that replication was not an end but a tool: a tool that could be employed for various purposes. This does not mean that their outcome was necessarily made up. Repetitions were done with a certain aim in mind, be this to confirm the original outcome or to reject them offering different data whenever the original study was not deemed convincing. We have reason to assume that when an experiment did not fulfill its aim, the

researcher did not feel inclined to publish the results (probably distrusting his method, subjects, and/or the workings of his instruments). At a later point, he might, nevertheless, suddenly use them to support his claim and findings.

To sum up, taking the historical replications together, we arrive at the following list of social functions and methodological purposes:

1. balancing out undesired variations (errors/confounding variables) in the measurements,
2. testing the stability of a finding (sometimes expressed in the form of a law or a statistical trend) using different experimental subjects, or repeating the experiment at a different place or time,
3. becoming acquainted with certain methods and gaining expertise within a scientific area,
4. training students who could then demonstrate their distinctive expertise and professional identity,

In these four cases, the replications aimed at the reproduction of exactly the same results as the previous (original) experiment which means that the first results are taken as authoritative guidance and calibration. Any variation could be worrisome, leading the experimenter to revise the apparatus and the way the experiment was done.

In other situations, in which an experiment was repeated, a difference with regard to the former results was acceptable and to a certain extent even expected. This would be the case for repetitions (replications) employed for the following purposes:

5. exploring the effect of some new variables (not varied or controlled in the original study),
6. standardized repetition was also seen as providing hints as to the particularities and types of an experimental subject’s mind,
7. cross-cultural appropriation: foreign scholars sought to learn how to do psychological experiments and to align their work with the scientific community, while pursuing their own religious and political agendas,

In the two cases of cross-cultural appropriations exposed, Gutberlet attempted to gain expertise by obtaining the same results while Navarro expected the Spanish mind to react differently to the stimuli.

Finally, yet another purpose for which an experiment was repeated was:

8. to challenge a psychological theory or empirical findings published by an authority in an impersonal way (letting the empirical findings, i.e., “nature” talk).

In the last case, the experimenter expected to arrive at different data or findings. I have argued that such a strategy

was used by Titchener to reject Wundt's theory while avoiding a personal confrontation. It was a strategy that could empower researchers from a weaker position or social status to question a well-established authority.

Finally, my historical examples can hardly give a definitive answer with regard to the broad question about the link between replicability and the wider context of academic life and modern society. Nevertheless, the present research seems to indicate some striking differences in the pace, time, and effort dedicated to replication. The reports of Fechner, Ebbinghaus and some of their followers capture their (and their experimental subjects') personal devotion and deep moral commitment to science as a way to "truth" that seemed to outweigh, at least to a certain extent, personal ambition. This can be seen in the numerous repetitions of their experiments as well as the careful selection of experimental subjects. While there could have been a file-drawer problem because researchers did not always publish immediately the results of all their experiments and replications, it is difficult to imagine early experimentalists such as Fechner, Ebbinghaus, Müller, Wundt, Titchener, or Calkins engaging in fraudulent practices such as p-hacking, even if they had had such a tool at hand, and even if it would have been beneficial for their academic careers. But whether this speculation is correct, is difficult to say.

### Declaration of Conflicting Interests

The author(s) declared no potential conflicts of interest with respect to the research, authorship, and/or publication of this article.

### Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

### ORCID.iD

Annette Mülberger <https://orcid.org/0000-0002-7260-9734>

### Notes

1. <https://osf.io/ezcuj/>
2. The file-drawer problem refers to the reluctance of researchers to submit their studies for publication unless they achieve a significant result. Romero (2019) expects around 5% of studies that are then published to show erroneous successes ("false positives"), while the "true negatives" often end up in the drawers of researchers' filing cabinets and are never published.
3. Schmidt (2009) introduced the distinction between direct and conceptual replication, which is still widely used although some authors have criticized or modified the distinction (Crandall and Sherman, 2016; Feest, 2019; Radder, 1992; Stroebe & Strack, 2014; Zwaan, Ets, Lucas, & Donnellan, 2018). In this paper, I follow Tweney's (2018) broad definition.
4. See also Radder (2003), who criticized Collins' work, emphasizing stability in experimentation, as well as the nonlocal meaning of experimental results. Collins' and Radder's clash reflects a much broader philosophical discussion that is still ongoing under headings such as the "realism/relativism dichotomy," "social constructivism," "nominalism," etc.
5. Fechner envisaged two kinds of psychophysics: external and internal. Here I deal only with the former, which became a fundamental approach in experimental psychology. For more information on his internal psychophysics, see Robinson (2010).
6. His philosophical world view called "day-view" (*Tagesansicht*) was inspired by romantic *Naturphilosophie* (see Heidelberg, 2004).
7. Gundlach (1993) summarized Fechner's position stating that ontologically mind and matter are identical but they differ perspectively. Heidelberg (2004) goes further, arguing that he defended a non-reductive neutral and materialist monism to which he added an "objective idealism." In the literature, Fechner's position is often presented vaguely as a "parallelist approach" or "double-aspect theory" (see Heidelberg, 2010, for more details).
8. His main psychophysical methods were: the method of the average error (Methode der mittleren Fehler), the method of adjustment (Methode der richtigen und falschen Fälle), and the method of limits (Methode der eben merklichen Unterschiede) (Fechner, 1860, p. 71).
9. This principle was later renamed the Principle of Indifference (see Stigler, 1986)
10. For more details on the mathematical formula and justification, see Gundlach (1993) and Heidelberg (2004).
11. Namely, the physicist Karl Hermann Knoblauch (1820–1895), the neurologist Richard Jung (1911–1986), and the physiologist Rudolf Heidenhain (1834–1897) who was also part of the family, married to Fanny Volkmann (1841–1867), daughter of physiologist Alfred Volkmann. For more information on Fechner's wide social circle, see Gundlach (1988).
12. See overview Köhler's (1886). Fechner made some corrections and wrote an extensive reply (Fechner, 1882).
13. This was facilitated by Fechner himself who had stated that psychophysics should only investigate the "phenomenal side of the material and the mental world" (Fechner, 1860, p. 8) and refrain from saying anything about the essences behind these phenomena. Similarly, Beiser (2020) argues that his panpsychist philosophy is more about the extent rather than the nature of mind.
14. Some biographical information about Gutberlet can be found here: [http://www.kathpedia.com/index.php?title=Constantin\\_Gutberlet](http://www.kathpedia.com/index.php?title=Constantin_Gutberlet)
15. For more information on the philosophical orientation of the Institution, see Jiménez-Landi (1996).
16. In Germany, they were mainly grouped around the two main journals of the time: the "Philosophische Studien" and the "Zeitschrift für Psychologie und Physiologie der Sinnesorgane."
17. Actually, Ebbinghaus himself presented his work this way (see: Preface, 1885, p. V). Jiménez-Landi (1996).
18. This qualification sounded much more negative in the 19th century and Ebbinghaus did not like the term for his approach (see his Abriss, Ebbinghaus, 1908).

19. He was also influenced by the work of the British associationists and the physiologists Hermann von Helmholtz and Ewald Hering. His approach is difficult to classify because he did not follow any specific psychological school or system, receiving vague descriptions as “methodology first” (van Rappard, 1987), “eclectic” (Boring, 1950), or “functionalist” (Caparrós & Anguera, 1986).
20. With regard to the origin of this study in Ebbinghaus’ dissertation, habilitation and the “Urmanuskript” of 1880, see Gundlach (1985).
21. Some recent examples of studies referring to Ebbinghaus’ forgetting curve and classical monograph, see Cyr and Hirst, 2019; Huang et al. (2021); and Otani et al. (2018).
22. Lifestyle refers to keeping a regular schedule for working, eating, and sleeping, avoiding, as much as possible, any disturbances such as travels or extreme emotional situations. See also Schaffer, 1988, on the importance of lifestyle for astronomical observations in the 18th century.
23. The reviewer was the psychologist Joseph Jacob, who published a short article on memory himself 2 years later; see Jacobs, (1887).
24. Georg Elias Müller (1850–1934) had been a student of Lotze and became his successor at the University of Göttingen, (Haupt, 2001). Following Boring (1935), Lotze’s support was due to his conviction that Müller possessed excellent qualities for doing experimental work in psychophysics. His co-experimenter, Friedrich Schumann (1863–1940), was at that time a doctoral candidate in physics who had chosen philosophy as his second subject. For more information on Schumann, see Lürer (2007) and Metzger (1940).
25. Nevertheless, in the end, they managed to finish all the series with five subjects: the experimental subjects were G.E. Müller and Schumann for each other’s experiments, together with the students Pilzecker, Hoffmann, and Hölzcke.
26. In the following years, Müller continued to work in the field of memory research making significant contributions. For example, with Pilzecker, he studied associations, showing that the speed with which they are reproduced reflects their strength (Müller & Pilzecker, 1900); and he later developed the Treffermethode (Müller, 1911, 1913, 1917).
27. See, for example, Anderson’s, 1985, and Kintsch’s 1985, critical appraisals in the 1980s.
28. For a more thorough response to Smedslund, see Teigen (1999); for a recent appraisal of Smedslund’s legacy in psychology see Lindstad, Stänicke, and Valsiner (2020).
29. My aim here is not to present all the replications and the reactions toward Ebbinghaus’ research. Hakes et al., 1964 and Young, et al., 1965 criticize some methodological shortcomings and for references including unsuccessful replications, see Murre and Dros, 2015.
30. The human relations within these laboratories were not the same (Kusch, 1999), but they all offered such practical laboratory training. At Leipzig, these were called “Psychologische Übungen,” and at the University of Cornell, students referred to them (though informally) as “Laboratory drill courses” (Boring, 1927, p. 497).
31. For information on the instruments used, see Wontorra (2013) and Haupt (2001). See also Gundlach (1996) on the symbolic role of the chronoscope.
32. Bringmann & Tweney, 1980; Danziger, 1990; Kusch, 1999. For a complete presentation of Wundt’s psychology, see Araujo, 2016. On the use of chronometry, see Schmidgen, 2004, 2014, and on introspection, see Danziger, 1980 and Feest, 2012,
33. For more detailed instructions on how introspective experiments work, he referred the reader in later editions of the Grundriss (Wundt, 1896/1913) to the textbook of his former student, Titchener (1900/1916).
34. The original term used at that time was “feeling” (Gefühl). It corresponds more to what we now understand as emotions, so this latter term is usually used in the secondary literature.
35. Sometimes the first pair is referred to as pleasantness–unpleasantness; and the second as excitement–tranquillization. For a discussion on Wundt’s terminology and the difficulty of English translation, see Titchener, 1908.
36. Called “objective symptoms” by Wundt and the method to register them: “Ausdrucksmethode.”
37. Given that the observations were made in the academic year 1897/98, Herr Watt might be referring to the American educational psychologist Guy Montrose Whipple (1876–1941) who joined Cornell University in 1898 and worked as an assistant in psychology until 1902. He received his PhD in 1900 under the supervision of Edward B. Titchener (Ruckmick, 1942)
38. In the eyes of other American psychologists such as Baldwin, this would still be considered too restricted (Danziger & Shermer, 1994).

## References

- Anderson, J. R. (1985). Ebbinghaus’s century. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 436–438. <https://doi.org/10.1037/0278-7393.11.3.436>
- Araujo, S. D. F. (2016). *Wundt and the philosophical foundations of psychology: A reappraisal*. Springer.
- Araujo, S. D. F., & Marcellos, C. F. (2017). From classicism and idealism to scientific naturalism: Titchener’s Oxford years and their impact upon his early intellectual development. *History of Psychology*, 20(2), 148–171. <https://doi.org/10.1037/hop0000059>
- Ash, M. G. (1980). Academic politics in the history of science: Experimental psychology in Germany, 1879–1941. *Central European History*, 13(3), 255–286. <https://doi.org/10.1017/S0008938900009602>
- Ash, M., & Geuter, U. (1985). *Geschichte der deutschen Psychologie im 20. Jahrhundert: Ein Überblick*. Westdeutscher Verlag.
- Bartlett, F. C. (1932). *Remembering: An experimental and social study*. In E. N. Zalta (Ed.), *The Stanford Encyclopedia of Philosophy (Spring 2020 Edition)*. Cambridge University. <https://plato.stanford.edu/archives/spr2020/entries/fechner/>
- Boring, E. G., & Edward, B. T. (1927). The American Journal of Psychology, 38(4), 489–506. <https://doi.org/10.1111/j.2044-8295.1928.tb00485.x>



- Boring, E. G. (1935). Georg Elias Müller: 1850-1934. *The American Journal of Psychology*, 47(2), 344–348.
- Boring, E. G. (1950). *A history of experimental psychology*. Appleton-Century.
- Bringmann, W. G., & Tweney, R. D. (1980). *Wundt studies: A centennial collection*. CJ Hogrefe.
- Buchner, E. F. (1900). Review of zur Kritik der Wundschen Gefühlslehre und Bemerkungen zur Theorie der Gefühle. *Psychological Review*, 7(1), 94–97. <https://doi.org/10.1037/h0067362>
- Buchwald, J. Z. (1989). *The rise of the wave theory of light: Optical theory and experiment in the early nineteenth century*. University of Chicago Press.
- Caparrós, H., & Anguera, B. (1986). Ebbinghaus y la tradición funcionalista. *Revista de Historia de la Psicología*, 7(4), 11–27.
- Carpintero, H. (2004). *Historia de la psicología en España*. Pirámide.
- Charle, C. (2004). *Patterns (chapt. 2)*. In W. Rüegg (Ed.), *A history of the university in Europe: Volume 3, universities in the nineteenth and early twentieth centuries (1800–1945)* (pp. 33–75). Cambridge University Press.
- Collins, H. (1985/1992). *Changing order: Replication and induction in scientific practice*. University of Chicago Press.
- Crandall, C. S., & Sherman, J. W. (2016). On the scientific superiority of conceptual replications for scientific progress. *Journal of Experimental Social Psychology*, 66, 93–99. <https://doi.org/10.1016/j.jesp.2015.10.002>.
- Cyr, T. G., & Hirst, W. (2019). Reflections on conversations and memory. *Topics in Cognitive Science*, 11(4), 831–837. <https://doi.org/10.1111/tops.12437>
- Danziger, K. (1979). The positivist repudiation of Wundt. *Journal of the History of the Behavioral Sciences*, 15(3), 205–230. [https://doi.org/10.1002/1520-6696\(197907\)15\(3\)<205::AID-JHBS15206696>3.0.CO;2-3](https://doi.org/10.1002/1520-6696(197907)15(3)<205::AID-JHBS15206696>3.0.CO;2-3)
- Danziger, K. (1980). The history of introspection reconsidered. *Journal of the History of the Behavioral Sciences*, 16(3), 241–262. [https://doi.org/10.1002/1520-6696\(198007\)16\(3\)<241::AID-JHBS160241>3.0.CO;2-3](https://doi.org/10.1002/1520-6696(198007)16(3)<241::AID-JHBS160241>3.0.CO;2-3)
- Danziger, K. (1987). *Hermann Ebbinghaus and the psychological experiment*. In W. Traxel (Ed.), *Ebbinghaus-Studien 2* (Passauer Schriften zur Psychologiegeschichte 5), pp. 217–222. Passavia.
- Danziger, K. (1990). *Constructing the subject*. Cambridge Univ. Press.
- Danziger, K., & Shermer, P. (1994). *The varieties of replication: A historical introduction*. In M. H. van IJzendoorn, & J. Valsiner (Eds.), *Reconstructing the mind: Replicability in research on human development* (pp. 17–36). Ablex Publishing Corp.
- Daston, L., & Galison, P. (2007). *Objectivity*. Zone Books.
- Derksen, M. (2021). *A menagerie of imposters and truth-tellers*. In S. Woolgar, E. Vogel, D. Moats, & C.-F. Helgesson (Eds.), *The Imposter as Social Theory: Thinking with Gatecrashers, Cheats and Charlatans* (pp. 53–76). Bristol University Press. <https://doi.org/10.2307/j.ctv1p6hphs.8>
- Draaisma, D. (1995). *Metaphors of memory: A history of ideas about the mind*. Cambridge University Press.
- Dror, O. (1998). *Modernity and the scientific study of emotions (1880-1950)*. Princeton University.
- Ebbinghaus, H. (1885). *Über das Gedächtnis (Untersuchungen zur experimentellen Psychologie)*. Duncker & Humblot.
- Ebbinghaus, H. (1908). *Abriss der Psychologie*. Veit.
- Fechner, G. T. (1860). *Elemente der Psychophysik (vol 1 and 2)*. Breitkopf & Härtel.
- Fechner, G. T. (1882). *Revision der Hauptpunkte der Psychophysik*. Breitkopf & Härtel.
- Feest, U. (2012). Introspection as a method and introspection as a feature of consciousness. *Inquiry*, 55(1), 1–16. <https://doi.org/10.1080/0020174X.2012.643619>
- Feest, U. (2019). Why replication is overrated. *Philosophy of Science*, 86(5), 895–905. <https://doi.org/10.1086/705451>
- Flis, I. (2019). Psychologists psychologizing scientific psychology: An epistemological reading of the replication crisis. *Theory & Psychology*, 29(2), 158–181. <https://doi.org/10.1177/0959354319835322>
- Frank, C. M., & Saxe, R. (2012). Teaching replication. *Perspectives on Psychological Science*, 7(6), 600–604. <https://doi.org/10.1177/1745691612460686>
- García Dauder, S. (2005). *Psicología y feminismo: Historia olvidada de mujeres pioneras en psicología*. Narcea Ediciones.
- Gigerenzer, G., Swijtink, Z., Porter, T., Daston, L., Beatty, J., & Kruger, L. (1989). *The empire of chance: How probability changed science and everyday life*. Cambridge University Press.
- Grower, B. (1997). *Scientific method: An historical and philosophical introduction*. Routledge.
- Gundlach, H. (1985). Das sogenannte Gedächtnis und das Unbewußte, oder Ebbinghaus' Versuch, Leibniz experimentell zu legitimieren. In D. Albert (Ed.), *Bericht über den 34. Kongreß der Deutschen Gesellschaft für Psychologie* (pp. 123–126). Hogrefe.
- Gundlach, H. (1988). Verwandte, Freunde, Vertraute. Die mathematisch-physische Abteilung der Königlich Sächsischen Gesellschaft der Wissenschaften zu Leipzig als Treibhaus der Psychophysik. In J. Brožek, & H. Gundlach (Eds.), *G. T. Fechner and Psychology. International Gustav Fechner Symposium Passau, 12 to 14 June 1987* (Passauer Schriften zur Psychologiegeschichte, 6) (pp. 87–102). Passavia.
- Gundlach, H. (1993). *Entstehung und Gegenstand der Psychophysik*. Springer-Verlag.
- Gundlach, H. (1996). The Hipp chronoscope as totem pole and the formation of a new tribe applied psychology, psychotechnics and rationality. *Teorie & Modelli*, 1(1), 65–85.
- Gutberlet, C. (1905). *Psychophysik: Historisch-kritische Studien über experimentelle Psychologie*. Von Kirchheim & Co.
- Hacking, I. (2014). *The taming of chance*. Cambridge University.
- Hakes, D. T., James, C. T., & Young, R. K. (1964). A re-examination of the Ebbinghaus derived-list paradigm. *Journal of Experimental Psychology*, 68(5), 508–514. <https://doi.org/10.1037/h0046779>
- Haseloff, O. W., & Jorswieck, E. (1970). *Psychologie des Lernens (Methoden, Ergebnisse, Anwendungen)*. De Gruyter.
- Haupt, E. J. (2001). Laboratories for experimental psychology: Göttingen's ascendancy over Leipzig in the 1890s. In R. W. Rieber, & D. K. Robinson (Eds.), *In: R. W. Rieber & D. K. Robinson, Wilhelm Wundt in History (The making of a scientific Psychology)* (pp. 205–250). Kluwer Academic / Plenum.

- Heidelberger, M. (2004). *Nature from within: Gustav Theodor Fechner and his psychophysical worldview*. University of Pittsburgh Press.
- Heidelberger, M. (2010). Functional relations and causality in Fechner and Mach. *Philosophical Psychology*, 23(2), 163–172. <https://doi.org/10.1080/09515081003727400>.
- Heller, O., Mack, W., & Seitz, J. (1991). Replikation der Ebbinghaus'schen Vergessenskurve mit der Ersparnisemethode: Das Behalten und Vergessen als Function der Zeit. *Zeitschrift für Psychologie*, 199(1), 3–18.
- Huang, Y., Fu, B., Lai, Y., & Yao, Y. (2021). Design and implementation of memory assistant based on Ebbinghaus forgetting Curve IOP conference series: Earth and Environmental Science. *IOP Conference Series: Earth and Environmental Science*, 687(1), 012187. <https://doi.org/10.1088/1755-1315/687/1/012187>.
- Irons, D. (1897). The nature of emotion. *Philosophical Review*, 7, 476–486.
- Iso-Ahola, S. E. (2020). Replication and the establishment of scientific truth. *Frontiers in Psychology*, 11, 2183. <https://doi.org/10.3389/fpsyg.2020.02183>.
- Jacobs, J. (1885). Reviewed Work(s): Ueber das Gedächtnis by Hermann Ebbinghaus. *Mind*, 10(39), 454–459. <https://doi.org/10.1093/mind/os-IX.34.296>.
- Jacobs, J. (1887). Experiments on “Prehension”. *Mind*, 12(45), 75–79. <https://doi.org/10.1093/mind/os-12.45.75>.
- Jiménez-Landi, A. (1996). *La Institución Libre de Enseñanza y su ambiente: Periodo de expansión influyente*. Edicions Universitat Barcelona.
- Kintsch, W. (1985). Reflections on Ebbinghaus. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 461–463. <https://doi.org/10.1037//0278-7393.11.3.461>
- Köhler, A. (1886). *Über die hauptsächlichlichen Versuche einer mathematischen Formulierung der psychophysischen Gesetze von Weber*. Philosophische Studien, 572–642.
- Kusch, M. (1999). *Psychological knowledge: A social history and philosophy*. Routledge.
- Lafuente, E. (1988). *Fechner in Spain: Some early responses to his psychophysics*. In J. Brožek, & H. Gundlach (Eds.) G.T. Fechner Psychology (Passauer Schriften zur Psychologiegeschichte, 6). Passavia.
- Latour, B., & Woolgar, S. (2013). *Laboratory life: The construction of scientific facts*. Princeton University Press.
- Laws, K. R. (2016). Psychology, replication & beyond. *BMC Psychology*, 4(1), 30–38. <https://doi.org/10.1186/s40359-016-0135-2>
- Leahey, T. H. (1981). The mistaken mirror: On Wundt's and Titchener's psychologies. *Journal of the History of the Behavioral Sciences*, 17(2), 273–282. <https://doi.org/10.1002/1520-6696>
- Lindstad, T. G., Stänicke, E., & Valsiner, J. (Eds.), (2020). *Respect for thought: Jan Smedslund's legacy for psychology*. Springer.
- Lüer, G. (2007). Georg Elias Müller (1850–1934): A founder of experimental memory research in psychology. *Cortex; a Journal Devoted To the Study of the Nervous System and Behavior*, 43(5), 579–582. [https://doi.org/10.1016/s0010-9452\(08\)70488-x](https://doi.org/10.1016/s0010-9452(08)70488-x)
- Makel, M. C., Plucker, J. A., & Hegarty, B. (2012). Replications in psychology research. *Perspectives on Psychological Science*, 7(6), 537–542. <https://doi.org/10.1177/1745691612460688>.
- Metzger, W. (1940). Friedrich Schumann, ein Nachruf. *Zeitschrift für Psychologie*, 148, 1–18.
- 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>.
- Mülberger, A. (2012). Wundt contested: The first crisis declaration in psychology. *Studies in History and Philosophy of Biological and Biomedical Sciences*, 43(2), 434–444. <https://doi.org/10.1016/j.shpsc.2011.11.002>
- Mülberger, A. (2017). Mental association: Testing individual differences before Binet. *Journal of the History of the Behavioral Sciences*, 53(2), 176–198. <https://onlinelibrary.wiley.com/doi/full/10.1002/jhbs.21850>.
- Mülberger, A., Gómez-González, C., Cañas, A. M., Cervantes, M., & Anglada-Llistosella, L. (2019). Testing the intelligence of Barcelona's schoolchildren in 1908. *Revista de Historia de la Psicología*, 40(1), 2–11. DOI:10.5093/rhp2019a1.
- Müller, G. E. (1911). Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufs. *Zeitschrift für Psychologie*, 5, 1–403.
- Müller, G. E. (1913). Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufs. *Zeitschrift für Psychologie*, 8, 1–567.
- Müller, G. E. (1917). Zur Analyse der Gedächtnistätigkeit und des Vorstellungsverlaufs. *Zeitschrift für Psychologie*, 9, 1–682.
- Müller, G. E., & Pilzecker, A. (1900). *Experimentelle Beiträge zur Lehre vom Gedächtnis*. J.A. Barth.
- Müller, G. E., & Schumann, F. (1894). Experimentelle Beiträge zur Untersuchung des Gedächtnisses (I & II). *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 6(28), 81257–190339. <https://doi.org/10.1093/mind/VII.28.588>.
- Murre, J. M. J., & Dros, J. (2015). Replication and analysis of Ebbinghaus' forgetting curve. *Plos One*, 10(7), Article e0120644. <https://doi.org/10.1371/journal.pone.0120644>
- Navarro Flores, M. (1915). *Manual de psicología experimental*. José Pijoán.
- Open Science Collaboration (2012). An open, large-scale, collaborative effort to estimate the reproducibility of psychological science. *Perspectives on Psychological Science: A Journal of the Association for Psychological Science*, 7(6), 657–660. <https://doi.org/10.1177/1745691612462588>
- Open Science Collaboration (2015). Psychology: Estimating the reproducibility of psychological science. *Science*, 349, aac4716. <https://science.sciencemag.org/content/349/6251/aac4716>.
- Otani, H., Schwartz, B. L., & Knoll, A. R. (2018). *Handbook of research methods in human memory* (pp. 1–18). Routledge.
- Pashler, H., & Wagenmakers, E. J. (2012). Editors' Introduction to the special section on replicability in psychological science. *Perspectives on Psychological Science*, 7(6), 52–530. <https://doi.org/10.1177/1745691612465253>
- Pettit, M. (2018). *Timeline: Replication in psychology: A historical perspective*. <https://psyborgs.github.io/projects/replication-in-psychology/>

- Porter, T. (1995). *Trust in numbers. The pursuit of objectivity in science and public life*. Princeton University Press.
- Porter, T. M. (2008). The objective self. *Victorian Studies*, 50(4), 641–647. [https://doi.org/10.1016/S0065-2601\(08\)60252-X](https://doi.org/10.1016/S0065-2601(08)60252-X)
- Radder, H. (1992). Experimental reproducibility and the experimenters' regress. *PSA: Proceedings of the Biennial Meeting of the Philosophy of Science Association, 1992*, 63–73. <https://doi.org/10.1002/9781118865064.ch4>.
- Radder, H. (2003). *The philosophy of scientific experimentation*. Pittsburgh Press.
- Rieber, R. W., & Robinson, D. K. (Eds.). (2001). *Wilhelm Wundt in history: The making of a scientific psychology*. Kluwer Academic/Plenum Publishers.
- Robinson, D. K. (2010). Fechner's "inner psychophysics." *History of Psychology*, 13(4), 424–433. <https://doi.org/10.1037/a0021641>
- Romero, F. (2017). Novelty versus replicability: Virtues and vices in the reward system of science. *Philosophy of Science*, 84(5), 1031–1043. <https://doi.org/10.1111/phc3.12633>.
- Romero, F. (2019). Philosophy of science and the replicability crisis. *Philosophy Compass*, 14(11), Article e12633. <https://doi.org/10.1111/pc3.12633>
- Ruckmick, C. A. (1942). Guy Montrose Whipple: 1876–1941. *The American Journal of Psychology*, 55(1), 132–134.
- Schaffer, S. (1988). Astronomers mark time: Discipline and the personal equation. *Science in Context*, 2(1), 115–145.
- Schickore, J. (2011). The significance of re-doing experiments: A contribution to historically informed methodology. *Erkenntnis*, 75(3), 325–347. <https://doi.org/10.1007/s10670-011-9332-9>
- Schmidgen, H. (2004). Die Geschwindigkeit von Gefühlen und Gedanken. *NTM International Journal of History and Ethics of Natural Sciences, Technology and Medicine*, 12(2), 100–115. <https://doi.org/10.1007/s00048-004-0190-2>.
- Schmidgen, H. (2014). *Hirn und Zeit: Die Geschichte eines Experiments 1800 - 1950*. Matthes & Seitz.
- Schmidt, S. (2009). Shall we really do it again? The powerful concept of replication is neglected in the social sciences. *Review of General Psychology*, 13(2), 90–100. <https://doi.org/10.1037/a0015108>.
- Shakow, D. (1930). Hermann Ebbinghaus. *The American Journal of Psychology*, 42(4), 505–518.
- Shapin, S., & Schaffer, S. (1985). *Leviathan and the air-pump*. Princeton University Press.
- Smedslund, J. (1987). Ebbinghaus, the illusionist: How psychology came to look like an experimental science. In W. Traxel (Ed.), *Ebbinghaus-Studien 2* (Passauer Schriften zur Psychologiegeschichte 5), pp. 225–240. Passavia.
- Stigler, S. M. (1986). *The history of statistics: The measurement of uncertainty before 1900*. Harvard University Press.
- Stroebe, W., Postmes, T., & Spears, R. (2012). Scientific misconduct and the myth of self-correction in science. *Perspectives on Psychological Science*, 7(6), 670–688. <https://doi.org/10.1177/1745691612460687>
- 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>
- Titchener, E. B. (1899). Zur Kritik der Wundt'schen Gefühlslehre. *Zeitschrift für Psychologie und Physiologie der Sinnesorgane*, 19(1), 321–326.
- Titchener, E. B. (1908). The tridimensional theory of feeling. *The American Journal of Psychology*, 19(2), 213–231. <https://doi.org/10.2307/1412760>
- Titchener, E. B. (1916). *A text-book of psychology*. Macmillan.
- Trautsholdt, M. (1883). Experimentelle Untersuchungen über die Association der Vorstellungen. *Philosophische Studien*, 1, 213–250.
- Traxel, W. (1987). *Ebbinghaus-Studien 2*. Passauer Schriften zur Psychologiegeschichte.
- Tulving, E. (1985). Ebbinghaus's memory: What did he learn and remember? *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 11(3), 485–490. <https://doi.org/10.1037/0278-7393.11.3.485>
- Tweney, R. (2018). *The origins of replication in science*. In J. F. Codell, & L. K. Hughes (Eds.), *Replication in the long nineteenth century: Re-makings and reproductions*. Cambridge. E-Book.
- Van Rappard, H. (1987). *Ebbinghaus' memory observed*. In W. Traxel (Ed.), *Ebbinghaus-Studien 2* (Passauer Schriften zur Psychologiegeschichte 5), pp. 35–46. Passavia.
- Wassmann, C. (2009). Physiological optics, cognition and emotion: a novel look at the early work of Wilhelm Wundt. *Journal of the History of Medicine and Allied Sciences*, 64(2), 213–249. <https://doi.org/10.1093/jhmas/jrn058>.
- 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>
- Witte, E. H., & Zenker, F. (2017). From discovery to justification: Outline of an ideal research program in empirical psychology. *Frontiers in Psychology*, 8, 1847. <https://doi.org/10.3389/fpsyg.2017.01847>.
- Wontorra, H. M. (2013). *Early apparatus-based experimental psychology, primarily at Wilhelm Wundt's Leipzig Institute*. In C. Pleh, L. Gurova, & L. Ropolyi (Eds.), *New perspectives on the history of cognitive science* (pp. 59–80). Akadémiai Kiadó.
- Wundt, W. (1888). Selbstbeobachtung und innere Wahrnehmung. *Philosophische Studien*, 4, 292–309.
- Wundt, W. (1896). *Grundriss der Psychologie*. Engelmann.
- Wundt, W. (1900). Bemerkungen zur Theorie der Gefühle. *Philosophische Studien*, 15, 149–218.
- Young, R. K., Hakes, D. T., & Hicks, R. Y. (1965). Effects of list length in the Ebbinghaus derived-list paradigm. *The Journal of Economic Perspectives: A Journal of the American Economic Association*, 70(3), 338–341. <https://doi.org/10.1037/h0022257>
- Zwaan, R. A., Etz, A., Lucas, R. E., & Donnellan, M. B. (2018). Improving social and behavioral science by making replication mainstream: A response to commentaries. *The Behavioral and Brain Sciences*, 41, 157–61. <https://doi.org/10.1017/S0140525X17001972>.