

“Repressed Memory” Makes No Sense

Felipe De Brigard

Department of Philosophy

Department of Psychology and Neuroscience

Duke Institute for Brain Sciences

Duke University

Address for correspondence:

201E West Duke Building

Box 90743

Duke University

Durham, NC 27708-0743

felipe.debrigard@duke.edu

“Repressed Memory” Makes No Sense

Abstract: The expression “repressed memory” was introduced over 100 years ago as a theoretical term purportedly referring to an unobservable psychological entity postulated by Freud’s seduction theory. That theory, however, and its hypothesized cognitive architecture, have been thoroughly debunked—yet the term “repressed memory” seems to remain. In this paper I offer a philosophical evaluation of the meaning of this theoretical term as well as an argument to question its scientific status by comparing it to other cases of theoretical terms that have either survived scientific change—such as “atom” or “gene”—or that have perished, such as “black bile”. Ultimately, I argue that “repressed memory” is more like “black bile” than “atom” or “gene” and, thus, recommend its demotion from our scientific vocabulary.

Keywords: Memory, Repression, Repressed Memories, Theoretical Terms, Theory Change.

“Repressed Memory” Makes No Sense

1. On humors and theoretical terms

Humor me for a second and let me start with a bit of historical fiction. It is the year 338 C.E. Octavia Cassia, the daughter of a rich merchant from Ephesus, has just moved with her husband to the old city of Byzantium—now Constantinople—partly for business but also to seek medical help. The sadness of a stillbirth a few months earlier continues to loom large, leaving her crestfallen and despondent. Her body, too, has weakened and turned pale. A flourishing city, brimming with academic and philosophical minds drawn from all over the known world, offers the most up-to-date medical knowledge, and with it a promise of recovery. A reputed physician is promptly called, and after a thorough examination, he confidently offers a diagnosis: Octavia’s melancholic demeanor and weakness is due to a humoral imbalance produced by an excess of black bile. The treatment is standard: a short bloodletting session that evening, with a follow-up on the next new moon, along with a healthy diet of cucumber, known to stabilize the secretions of the spleen.

Both the diagnosis and treatment followed directly from the prevailing medical doctrine of the time. Indeed, the *humoral theory*, or the view according to which human beings are composed of four elementary fluids or “humors”, constituted the theoretical backbone of our medical practice for centuries. Although typically traced back to *The Nature of Man*—a Hippocratic treatise compiled by Hippocrates’ student and son-in-law, Polybus, in the 5th century B.C.E.—the truth is that, back then, there were several humoral theories in the offing, each proposing slight variations on the nature of the four basic fluids (Jouanna, 2012). For various reasons, Hippocrates’ set of four humors—blood, phlegm, yellow bile, and black bile—prevailed, reaching wide acceptance by the

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

time Galen's *Commentary* on the Hippocratic treatise saw the light in the 2nd-century C.E. Our best historical estimates indicate that the humoral theory continued to reign well into the Renaissance, even managing to show up in several of Shakespeare's plays. Thanks to numerous developments in the 16th century—including Paracelsus' new iatrochemical approach to health as well as Vesalius' unrelenting corrections of Galen's anatomical observations and Harvey's discovery of the circulation of blood—the humoral theory eventually fell in disrepute. By the 17th century it was all but replaced by a mechanistic view of the human body (Bos, 2009).

Humorism enjoyed an extraordinarily long life, and, in many ways, it embodied central characteristics of scientific theories. One such feature, critical for our present purposes, was its appeal to unobservable entities to explain puzzling observable phenomena. People who overproduce mucus and other secretions, for example, also tend to appear lethargic and apathetic. According to humorism, both behaviors were jointly explained by a common unobservable cause: excess of phlegm (hence the word “phlegmatic”). In fact, of all the humors, blood was the only one visible to the naked eye. All other humors were unobservable, although they could have unique physical effects, such as the dark color in certain kinds of vomit, or the presence of “lees of blood”, which is what Galen called the sediment found at the bottom of glass vials in which blood has been left sitting, unperturbed, for a while (Saher et al., 2021). Thus, terms like “phlegm” or “black bile” featured in the humoral theory as theoretical terms: referential expressions employed in scientific theories to denote non-observational entities and whose meaning is determined by the theory alone (Andreas, 2021). Scholars disagree as to why exactly those four humors were chosen. Some argue that they were associated to the four seasons: blood for the spring, yellow bile for the summer, black bile for the fall, and phlegm for the winter. Others suggest that they were taken to parallel the four stages of the human life (infancy, youth, maturity, old-age), or even ancient theories of

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

the four basic elements (fire, air, earth, and water). In any case, all accounts agree that the existence of such humors was assumed on theoretical, as opposed to observational, grounds.

Nowadays, of course, Octavia's diagnosis and proposed treatment sound ludicrous. This isn't simply because the physician's diagnosis is false, given that there was no black bile in Octavia's body. If you seek to state a matter of fact (as opposed to, say, fiction), and your statement referentially employs a theoretical term purportedly denoting an unobservable entity, then the non-existence of the entity is likely sufficient to render the statement false. As a comparison, consider the following case. Circa 1980, an astronomer, upon observing an anomaly in the orbit of a distant planet, asserts that "the planet deviated from its expected position because there is a nearby black hole exerting strong gravitational force". Here, "black hole" is a theoretical term purportedly referring to a region of spacetime whose gravitational force is so strong that light cannot bounce out of it, rendering it—at least back in 1980—unobservable. As it turns out, though, there was no such black hole, so the assertion is false. But notice that the case of the astronomer seems importantly different from that of the Byzantine physician. Given the nature of planetary movements and their gravitational interactions with asteroids, the astronomer's claim, while false, could have been true in a nearby possible world in which our physical and cosmological laws are kept constant. Indeed, we could even learn something from the astronomer's false statement, namely that had a black hole been in the location the astronomer thought it was, then, *ceteris paribus*, it would have altered the asteroid's orbit in just such a way. By contrast, in the case of Octavia's diagnosis, there is simply no nearby possible world in which the physician's statement could have been true at the same time as our biological and physical laws. For that statement to be true, not only our biology but likely also our physics would have to be entirely different. One gets

the impression that the physician's assertion, to use Pauli's famous quip, "is not only not right; it is not even false."

Two insights from work in the semantics of theoretical terms can help us to better capture the difference between the cases of the astronomer and the Byzantine physician. The first insight comes from Frege (1892) and Husserl (1900), who argued that referential terms have both *sense* and *reference*—or 'connotation' and 'denotation', for Husserl. Roughly, the sense of a term is the way in which we conceive of or understand it, whereas its reference is that which the term denotes. For someone unaware of his secret identity, the terms "Superman" and "Clark Kent" express different senses, even though they both have the same referent: Kal-El. The second insight comes from what sometimes is called *contextual semantic theory*, according to which the meaning of a scientific term—whether observational or non-observational—depends on how it is incorporated within the theory (e.g., Poincaré, 1902; Duhem, 1906; Quine, 1951; Feyerabend, 1962). The sense of a scientific term is therefore going to depend upon a complex web of empirically corroborated statements, observations, predictions, and validated practices, all of which specify the way in which its referent is to be determined. Putting these two insights together, then, we get that the statements of both the astronomer and the Byzantine physician are false insofar as they lack a referent—they are false in the same way in which, say, "The present king of France is bald" is false (Russell, 1905). However, in the case of the astronomer's assertion, the sense of the theoretical term "black hole" is given by an extraordinarily successful theory, replete with fruitful predictions, satisfactory explanations, and confirmed hypotheses. The astronomer's assertion was false because it lacked a referent, but it didn't lack sense. By contrast, the theoretical term "black bile" employed by the Byzantine physician is now isolated from any successful biological or physical theory; the humoral theory has been debunked—it has been shown to fail predictively

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

and explanatorily, and to offer no testable hypotheses. As such, when we now read the diagnosis of the Byzantine physician, we deem it false not only because it lacks a referent, but also because it makes no sense. Given our current successful and fruitful biological and physical theories, it is unclear how to go about determining the referent of such a theoretical term. We would have to unknow too much to get close to grasping what it could have meant.

2. On the origins of “repressed memory”

I used the example of “black bile” within the humoral theory because my plan is to argue now, by analogy, that the same goes for the theoretical term “repressed memory”. Tradition tells us that the term was first introduced by Freud in 1896, in a lecture that was later published with the title “On the etiology of hysteria”—although something very close to that notion was already present in his work with Breuer (Breuer and Freud, 1895).¹ As with most theoretical terms, the expression “repressed memory” was introduced both to help to explain puzzling observable phenomena and to refer to an unobservable entity. The puzzling observable phenomena it sought to explain was a constellation of behaviors considered to be symptoms of hysteria.² Its alleged symptomatology was truly a menagerie of conducts that included shortness of breath, paralysis, amnesia, vomiting, selective deafness, repetitive movements, and infertility, among many others. The notion of “repressed memory,” then, offered an attractive, unifying causal factor that could help to explain these disparate behaviors.

¹ According to Erdelyi (2006), the term “repression” was first employed by Herbart in 1824, in work that was known by Freud, and which seems to have influenced him. For Herbart, however, “repression” simply referred to the act of an idea inhibiting another idea—perhaps similar to the way in which “blocking” works today—and thus was not imbued with the causal/functional role Freud gave to the term. Thus, despite this historical precedent, it still holds that the first use of the term “repressed memory” qua theoretical notion, was Freud’s—and Erdelyi (2006) agrees.

² For most history, hysteria was exclusively diagnosed to women. Indeed, the term comes from the Latin “*hysteria*”, or “uterus”, as it was thought that the behaviors were produced by women experiencing a “wandering uterus”. Freud is credited with being the first to diagnose hysteria to men, but this is incorrect: years before, Jean-Martin Charcot (1887) had characterized men with the alleged disorder, which he thought of as a disease of the central nervous system.

The nature of this putative unobservable factor—these repressed memories—was characterized in terms of the causal/functional role they played within a hypothesized cognitive architecture. Sketched in his “seduction theory”, Freud conceived of the mind as being composed of the conscious and the unconscious. Accordingly, when we experience events, we form memories that get stored in our—apparently limitless—unconscious. According to the theory, every memory has a certain level of “traumatic force”, which is manifested in their tendency to move from the unconscious to the conscious. Most of our memories presumably have little traumatic force, and are left behind in the unconscious, forgotten. Some, by contrast, are retrieved but cause no bizarre behavioral manifestations, as their traumatic force is minimal. But there is a subset of those memories, high in traumatic force, which the mind seeks to keep in the unconscious, as they would cause havoc were they to reach the conscious. To help to control these memories, the mind makes use of a defense mechanism dubbed “repression” whose function is to keep memories with high traumatic force at bay. Thus, the term “repressed memories” was introduced in reference to memories with high traumatic force that are kept at bay in the unconscious by the gating operation of repression.

When repression needs to relieve pressure from the unconscious, which is not infrequently, it loosens its grip a bit, letting memories from the unconscious filter through from time to time and manifest in the conscious, often in the form of dreams and slips of the tongue. However, when the memories have quite a bit of traumatic force, repression can no longer retain them, and they tend to manifest in all sorts of pathological behaviors, such as those that comprise the symptomatology of hysteria. Moreover, according to the seduction theory, the repressed memories that make their way to the conscious and cause individuals to behave hysterically, are all sexual in nature; specifically, they are all memories of sexual abuse experienced in early childhood but kept

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

unconscious by the repression until they start to manifest in pathological behaviors. Intriguingly, repressed memories could manifest through the conscious without their content—i.e., what they were about—being made consciously available to the individual. Such content could thankfully become available to the subject by the grace of psychoanalytic therapy.

Freud abandoned some of the main tenets of the seduction theory shortly after he proposed it, in part—as he confided to his friend Wilhelm Fliess on September 21, 1897—because he felt that the proportion of individuals exhibiting hysterical behaviors likely outnumbered the incidence of infantile sexual abuse. But he soon accommodated this apparent counterevidence by claiming instead that the memories from the unconscious, which can lead to pathological behaviors, need *not* be of experienced events that occurred in one’s childhood: they can actually come from before one was born. Indeed, Freud tells us that, for a child, to be born is already profoundly traumatic, and that the life of every infant is, from its very onset, suffused with sexual frustration and distress. Everyone can develop pathological behaviors because everyone, by the sheer fact of being human, must repress an enormous amount of sexual trauma. What should surprise us, then, is that the proportion of hysterical behaviors is not higher.

Freud’s model of the architecture of mind as composed of the unconscious and the conscious, as well as this mysterious mechanism of defense he called “repression” was, well, seductive.³ As a result, his “topographical model”, as it is sometimes called, became profoundly influential during most of the 20th century, and in some areas of the world—particularly in Europe and Latin America—continues to be popular among certain psychiatrists. Nevertheless, as a theory of our cognitive architecture, Freud’s topographical model has been thoroughly debunked. Almost

³ Freud modified his basic model a few times. He introduced, for instance, the notion of “preconscious” to refer to the place in which memories that are not conscious but not repressed, reside (Freud, 1900). A more substantial revision occurred a couple of decades later, when he overwrote his tri-partite distinction with that of the Ego, Id and Superego, which do not quite map onto his original “topographical” system, from which the term “repressed memory” emerged.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

every aspect of his theory has been criticized from basically every angle. Early on, scholars argued that some of its key notions were incoherent and internally contradictory (Wells, 1913) as well as false or, at best, metaphorical (Haberman, 1914). Freud's view of the mind was famously used by Popper as a paradigm of a non-scientific theory, as its main claims were unfalsifiable and offered no testable predictions (Popper, 1962; 1974). More devastatingly, perhaps, Grünbaum (1984) argued that even the apparent therapeutic success of psychotherapy could not be validly leveraged in support of the theoretical assumptions of Freud's view.

More circumscribed criticisms against the existence of repressed memories and of the alleged cognitive architecture in which they could occur, have also filled the literature. For instance, there is little, if any, experimental evidence to the effect that "the unconscious" and "the conscious" name independent cognitive systems. To my knowledge, no study has shown a single, let alone a double, dissociation between these two putative systems, and no experimental task that I know of seems to successfully manipulate repression as a gateway for memories between the two. Even though some have tried to link existent evidence of unconscious processes to the Freudian notion of unconscious (e.g., Kihlstrom, 1990, for a review), the fact remains that such results are entirely interpretable independently of any Freudian doctrine. Moreover, extant studies that supposedly demonstrate the existence of repressed memories have also been widely discredited, either because they were methodologically faulty (Kihlstrom, 1996), or because they could not rule out that the alleged recovered memories were not iatrogenic, as suggested by the many demonstrations of false implanted memories in the lab and during therapy (Loftus, 1994). In sum, there is absolutely no evidence to the effect that our mind contains a repository of memories, each of which has a certain level of "traumatic force", eagerly inching toward consciousness, yet kept at bay by a "defense mechanism" whose guard is only let down when you

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

drink, dream, slip your tongue, or, obviously, reach transference with a psychoanalyst who, for an appreciable fee, can help you control your repression.

As a theory of the architecture of the human mind, Freud's topographical model is as inaccurate as the humoral theory is of the nature of the human body. Consequently, by parity of argument, I surmise that when the term "repressed memory" is referentially used in an explanation to denote the alleged unobservable factor purportedly responsible for certain behaviors, including the sudden recovery of a vividly disturbing recollection, its semantic status is no different than that of "black bile" when employed by the Byzantine physician. In both cases the explanatory statement would be false, not only because the theoretical term lacks a referent, but also because it makes no sense.

3. On dissociative amnesia, blocked memories, and intentional forgetting.

At this point, someone may argue that although the term "repressed memory" may have lost its original intended meaning, due in no small part to the failure of the Freudian topographical model, the term nonetheless has acquired a new sense in contemporary science. There are two ways in which one could interpret this objection. On the one hand, it could be interpreted as suggesting that the term "repressed memory" makes sense today, in part, because the unobservable entity it putatively refers to has survived theoretical change. On the other hand, it could be interpreted as suggesting that even though its putative referent has *not* survived theoretical change, the term itself has acquired a new sense in contemporary science. I explore both objections in turn.

3.1. Same referent, different terms

Scientific terms surviving theoretical change are not uncommon. Theoretical terms such as “atom” and “gene”, for instance, originated within physical and biological theories that have changed, and in certain aspects bear little resemblance to contemporary physics and biology. Nevertheless, the terms persist, roughly referring to—many argue—the same unobservable entity that they were meant to denote. Could this also be the case, then, with the term “repressed memory”? In fact, in the last couple of decades, at least three expressions have been suggested as tantamount to the theoretical term “repressed memory,” namely “dissociative amnesia,” “blocked memories,” and “intentionally forgotten memories.” Might it be that the notion expressed by the initial term persists, yet disguised under a different name?

To evaluate this possibility, it is important to understand, first, why some theoretical terms, such as “atom” or “gene”, managed to survive theoretical change, while others, such as “black bile”, perished. One promising strategy is to make use of a causal/functional approach for reference fixing (Psillos, 1999), according to which

- (1) a term t refers to an entity x if and only if x satisfies the core causal/functional description associated with t , and
- (2) two terms, t and t^* , refer to the same x if and only if
 - (a) their putative referents play the same causal/functional role with respect to a relevant set of observable phenomena; and
 - (b) the core causal/functional description of t^* shares the same constitutive characteristics of the core causal/functional description of t .

To illustrate, suppose that the core causal/functional description of “atom” essentially involves something like “being the elementary particle that is constitutive of all matter”. Then, it can be argued that although the theoretical characterization of atoms has changed through time, so that

the sense under which Dalton understood “atom” in 1817 is different from that under which, say, Heisenberg understood it in 1927, they nonetheless share a core causal/functional description. Likewise, one could argue that if the core causal/functional description of “hereditary factors” (Mendel, 1865) essentially involves something like “being the material unit of inheritance,” then what Johannsen (1909) meant by “gene” is continuous, not only with Mendel’s original notion, but also with our sense of “gene”, even though the theories in which both terms were coined are importantly different from our contemporary understanding of genetics.

Alas, I do not think that the case of “repressed memories” is analogous to that of “atom” or “gene,” for two reasons. First, because none of the three candidate replacements share the same constitutive characteristics of the core causal/functional description under which the notion of repressed memory, and its corresponding theoretical term, were introduced. And second, because the sets of observable phenomena allegedly explained by each of these alternative theoretical terms, are non-coextensive.⁴ To start, consider the notion of “dissociative amnesia”. Previously known as “psychogenic amnesia”, the term was introduced in the DSM-5 (American Psychiatric Association, 2013) to refer to very rare cases of individuals presenting severe inability to recall past personal information, with the caveat that the deficit cannot be explained by normal forgetting or by otherwise detectable neurological damage—what once upon a time was known as not having “organic origin” (hence “psychogenic”). The problem is that the observable phenomena dissociative amnesia is supposed to explain is both narrower and wider than the set of observable

⁴ To clarify this point: Suppose that a theoretical term, t , is introduced to explain a set of observable phenomena, $A = \{a, b, c, d, e\}$. Then, if someone were to argue that another theoretical term, t^* , is co-referential with t , then it should be the case that the set of observable phenomena explained by t^* is co-extensive with the set of observable phenomena explained by t . To make this point even clearer, suppose that t^* accounts for the set of observable phenomena $B = \{b, c, d, f\}$. As such, although there is an overlapping set, $A \cap B = \{b, c, d\}$, there are observable phenomena t accounts for that t^* does not, $\{a, e\}$, as well as observable phenomena that t^* accounts for that t does not, $\{f\}$. Consequently, even if overlapping, when the sets of observable phenomena two theoretical terms are thought to account for are non-coextensive, then there is reason to believe that both terms are not co-referential.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

phenomena repressed memories were supposed to account for. It is narrower in the sense that, as mentioned before, repressed memories were called to explain a wide variety of behaviors, almost none of which are included in the symptomatology listed for dissociative amnesia. And it is wider in the sense that some behaviors, indicative of dissociative amnesia, do not conform to the way in which repressed memories are supposed to operate. For instance, one could be diagnosed with dissociative amnesia—which, for the record, is an extremely rare diagnosis—when presenting selective global amnesia, whereby one is incapable of recalling information from an entire time period. But this is not the way in which repressed memories were thought to work. Recall that, according to Freud, each particular memory in the unconscious has a certain “traumatic force”, and that it is this individual memory that can cause havoc in the conscious. Repressed memories were not blocked *en masse*, for the mnemonic content of the individual memory is critical to explain the relevant abnormal behavior.

To be fair, some cases of dissociative amnesia involve selective memory loss of individual events, but then again this is the only point of intersection between two otherwise disparate set of observable phenomena (footnote 4). Moreover, the causal/functional descriptions of both notions are also different. As mentioned, a “repressed memory” was supposed to refer to an individual memory with high traumatic force, due to the sexual nature of its mnemonic content, which has been kept at bay in the unconscious by repression. By contrast, the causes of dissociative amnesia—which are largely unclear (Staniloiu and Markowitsch, 2014)—do not involve repression but, instead, external factors leading to the individual’s overwhelming stress at the time of encoding and retrieval. Additionally, the relevant memories affected by dissociative amnesia need not be sexual in nature. Thus, although there is some overlap in the set of observable behaviors “repressed memory” and “dissociative amnesia” were supposed to account for, as well

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

as a family resemblance between their causal/functional characterizations, the fact remains that both conditions (1) and (2) are violated, leaving us with little reason to believe these two terms are co-referential.

Similar considerations apply to the notion of “blocked memories” and to memories that have been “intentionally forgotten”. In a recent review paper, for instance, Smith and Beda (2020) criticize those who denied “the existence and reality of recovered memories” during the so-called “memory wars” and argue that there is a legitimate sense in which memories can be repressed and later recovered. But the kind of phenomena they have in mind involves small and transient effects such as tip-of-the-tongue or drop-out induced forgetting, in which a sub-set of encoded target items are dropped from further exposure to maximize distractor interference, effectively “blocking” them at retrieval. Once again, these kinds of phenomena are not co-extensive with the set of phenomena repressed memories were supposed to explain, and the causal/functional description of their putative mechanisms are entirely different. In particular, Smith and Beda (2020) advocate for a two-process model that involves the dominance of distractors and targets, and makes no reference whatsoever to notions like traumatic force, unconscious, or repression.

The same argument applies to recent attempts to use experimental results showing “intentional forgetting” as evidence for repressed memories. A typical intentional forgetting study involves variations of the ‘think/no-think’ paradigm, whereby participants are presented with a cue and an associated item, say an unrelated word, and then, upon presentation of the cue, they are instructed to either think or not think about the associated item. Then, in a subsequent memory test, participants are presented again with the cues and are asked to recall all the associated items. Intentional forgetting is thus measured as a proportion of the remembered associated items in the ‘think’ relative to the ‘no-think’ condition. But, then again, this is very different from the

behavioral phenomena repressed memories were supposed to account for. For one, the measure of intentional forgetting is usually a reduction in the rate of retrieval across several items, not evidence for the intentional forgetting of one specific item. Indeed, if the experiment is well designed, ideally you would control for item specificity, as you want to show the effect irrespective of the precise forgotten item—which is exactly the opposite of what you’d expect for a particular memory that is, allegedly, repressed. Moreover, as Otgaar et al (2020) reminds us, there is no evidence that you can use the think/no-think paradigm to completely obliterate a past autobiographical memory. Even nice demonstrations of the use of this paradigm with autobiographical memory (e.g., Noreen and MacLeod, 2013) show retrieval way above chance, and the effect is just about a 10% reduction relative to baseline. More importantly, there is no evidence whatsoever to the effect that those few memories forgotten during the memory test are lost forever, or even for any prologued period of time after the participants leave the lab. And finally, there is evidence that insofar as the intentional forgetting effect is real, it tends to be rather small, unstable, and hard to replicate (Wessel et al., 2020). In sum, “intentional forgetting” is a theoretical term used to explain a very different range of observable phenomena, and via a very different causal/functional description, than the phenomena allegedly covered by the term “repressed memory”.

3.2. Same term, different referent

The above objection can also be interpreted as suggesting that the term “repressed memory” persists today, not because its putative referent has survived theoretical change, but because it has been endowed with new meaning, grounded not in the feeble soil of Freud’s seduction theory, but in the solid foundation of contemporary psychological science. According to

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

this view, then, the sense of the term “repressed memory” today has very little, if anything, to do with the sense Freud’s sought to give to the term 100 years ago. As a result, Freud’s “repressed memory” and today’s “repressed memory” are definitively not coreferential; at best, they are distant cousins, at worst, they are merely homonyms.

The force of this objection comes from the observation that, since its inception, several researchers have sought to save the notion of repressed memory by proposing different characterizations for the term, all of which deviate in important ways from Freud’s original formulation. Anna Freud (1936), for instance, proposed to assimilate the notion of repression to the unconscious alone, as opposed to leaving it as a putatively independent mechanism in-between the conscious and the unconscious. She further suggested that *suppression* was its conscious counterpart, allowing now the possibility of two defense mechanisms—repression and suppression—operating over two allegedly distinct kinds of ideas. Unfortunately, this proposal introduced more trouble than it solved, not only because it seems to directly contradict Freud’s own views (Erdelyi, 2006), but also because it doubled the number of alleged psychological defense mechanisms whose ontological status was problematic to begin with. Moreover, her proposal was interpreted differently by the psychoanalytic establishment in the years to come, with some claiming that repression was confined to the realm of the unconscious and the involuntary, while others argued that suppression was just conscious repression and, thus, that it could happen voluntarily and with an individual’s full awareness (Cramer, 1998). Indeed, much of the recent debate around the term “repressed memory” revolves around disagreements on precisely these further characterizations, with some arguing that repressed memories are unconscious, other that they are conscious, some that they are voluntary, and some that they are not (e.g., Brewin et al., 2019; Otgaar et al., 2020).

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

Space limit prevents me from discussing the particularities of each one of these proposals, but there is a general point that can be made: as far as I can tell, all contemporary characterizations of the term “repressed memory” typically make use of additional theoretical notions that tend to be vague or, at least, underdefined. Take, for instance, the claim that memory repression can be thought of as a voluntary or deliberate strategy to prevent the retrieval of an unwanted memory (e.g., Brewin et al., 2019). Seems simple enough, until one wonders what “voluntary” may mean here. One possibility is that “voluntary” means something akin to a *synchronic* mental act, in which one tries not to think about a particular memory when one feels that a cue is about to trigger it. But then it would be strange to call that process “memory repression” as it seems to be just an instance of a more general strategy for thought suppression via attentional disengagement or distraction. Are words that we try to keep at bay while playing Taboo repressed memories? A synchronic interpretation of “voluntary” as a mental act would not rule them out. But another possibility is to interpret “voluntary” as meaning something like a *diachronic* strategy, in which one purposefully avoids contexts that could remind us of a negative event (Bermudez et al, 2021). For instance, I can voluntarily choose a daily route that makes it unlikely I will encounter cues that would normally trigger a negative memory, which in turn reduces the probability of retrieving it and the chances that I will forget it. Does that count as voluntary memory repression? If so, then likely some of the mechanisms involved are going to be unconscious and not dissimilar from those engaged in regular forgetting. Moreover, similar considerations apply to attempts to rescue the scientific standing of “repressed memory” on account that people commonly use “unconscious/conscious memory repression” in everyday parlance (e.g., Brewin et al., 2019; Otgaar et al., 2020). Alas, both “conscious” and “unconscious” are profoundly complex and ambiguous expressions,

and there is plenty of evidence that both experts (Lycan, 1996) and non-experts alike use them inconsistently (De Brigard, 2010).

In sum, the idea that the term “repressed memory” has acquired a new sense within a contemporary scientific theory is problematic. The few attempts to characterize the notion in more contemporary terms tend to involve expressions that are regularly vague or underdefined, and often a minimal scrutiny suffices to make us doubt whether such a notion could denote a distinct psychological entity. I know that psychologists often make use of the notion of “operationalization” to argue for the existence of a putative psychological process, but even the best characterizations of “repressed memory” are hardly operationalizations. Recall that the term “operationalization” is derived from an approach in physics, often called “operationalism”, in which a process that is not directly observable is measured indirectly (Bridgman, 1927). The theoretical term with which such a process is denoted is then definitionally associated with a particular way of measuring it. This is not the case with the characterizations of “repressed memory” available today, as they are not tied to particular measures, let alone experimental manipulations of any sort. And even if they were, there would still be a further question as to whether they name real entities. The operationalization of an unobservable term is not a sufficient condition for the existence of its referent (De Brigard, 2020). Today, the term “repressed memory” is far from having received an operational definition, and farther still for it being accepted as naming something real.⁵

⁵ One last possible interpretation of the objection, which may not deserve its own sub-section but certainly its own footnote, is to think that the notion of repressed memory persists today but with a different name *and* a different referent. A reviewer pointed out that this may be the case with the expression “dissociative amnesia” which, for some, is synonymous with “repressed memory”, even though they are not co-referential (Pope et al, 2023). But this may as well be an argument for *eliminating* the term “repressed memory” instead. Here is an analogy. The existence of aether as a propagation medium for light was accepted by many, including Newton, as the best explanation for a number of observable phenomena. By the time Einstein’s general relativity theory came along, the referent of the term “aether” was no longer thought to exist, and instead the phenomena it was employed to account for was instead explained by gravitational forces and the curvatures of space-time. However, it would seem wrong today to say that “gravitational

4. Conclusion

The purpose of this paper, which can be thought of as comprising two parts, is to offer a philosophical evaluation of the semantics of the term “repressed memory”. The first part, developed in sections 1 and 2, is an argument by analogy. One analogue is a purportedly explanatory statement that includes the theoretical term “black bile”. I argued that such a statement is false, not only because the term “black bile” has no referent, but also because it has no sense. This is because the theory in which the term was coined, and which was used to make sense to the term “black bile”, is now utterly debunked. The other analogue is a purportedly explanatory statement that includes the term “repressed memory”. I argue, by analogy, that such a statement is false for exactly the same reason: it lacks both referent and sense, since the theory in which the term was coined, and which was used to make sense to the term, is also utterly debunked. The second part, which occurs in section 3, explores two possible objections to my argument. The first objection is that, unlike “black bile”, the term “repressed memory” is akin to other theoretical terms—such as “atom” or “gene”—that have managed to survive theoretical change, albeit with a different name. I offer three candidate terms—i.e., “dissociative amnesia”, “blocked memories” and “intentional forgetting”—but argue that their meanings are not equivalent to that of “repressed memory”, so they don’t show that the term has survived theoretical change. Finally, the second objection is that the term “repressed memory” is alive and well, not because it has survived theoretical change, but because researchers characterize it anew, endowing it with a novel sense,

forces” is basically synonymous with “aether”. If the term “dissociative amnesia” does, in fact, refer to a real psychological entity, and such an entity is not the same as what the term “repressed memory” was introduced to refer to, then why keep the old, outdated term around? Of course, a different question is whether the term “dissociative amnesia” refers to a real psychological entity at all, or even a useful psychological construct, which remains very controversial (McNally, 2023; Pope et al., 2022).

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

grounded in currently accepted science. Unfortunately, I argue that this strategy promises more than it delivers, for the few available contemporary characterizations of the term “repressed memory” are vague, incomplete, or insufficient to legitimize its scientific credentials. Let me be clear, to conclude, that in no way am I saying that people cannot forget traumatic events or that sometimes long forgotten episodes cannot be brought back to mind, unexpectedly. These are all-too-common experiences whose reality is not questioned. What I am challenging is that these occurrences can be explained in terms of so-called “repressed memory”, for when this theoretical term features among the explananda, it not only lacks a referent: it makes no sense. And the fact that people may be employing it metaphorically or figuratively is no reason to keep it in our scientific practice.

Acknowledgments: Many thanks to Jamin Asay, Caleb Hazelwood, Anna Smith, Wayne Norman, Katherine Brading and Andrew Janiak for their help while writing this paper.

References

American Psychiatric Association. (2013). Diagnostic and statistical manual of mental disorders (5th ed.). <https://doi.org/10.1176/appi.books.9780890425596>

Andreas, H. (2021). Theoretical terms in science. *The Stanford Encyclopedia of Philosophy*.

Bermúdez, J.P., Murray, S., Chartrand, L. and Barbosa, S. (2023). What’s inside is all that counts? The contours of everyday thinking about self-control. *Rev.Phil.Psych.* 14, 33–55.

Bos, J. (2009). The rise and decline of character: humoral psychology in ancient and early modern medical theory. *History of the Human Sciences*, 22(3), 29-50.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

Bridgman, P. W. (1927). *The logic of modern physics* (Vol. 3). Macmillan. Breuer, J. and Freud, S. (1895). *Studies on hysteria*.

Brewin, C. R., Li, H., Ntarantana, V., Unsworth, C., & McNeilis, J. (2019). Is the public understanding of memory prone to widespread “myths”? *Journal of Experimental Psychology: General*, 148(12), 2245.

Cramer, P. (1998). Coping and defense mechanisms: What's the difference?. *Journal of personality*, 66(6), 919-946.

De Brigard, F. (2010). Attention, Consciousness, and Commonsense. *Journal of Consciousness Studies*. 17 (9-10): 189-201.

De Brigard, F. (2020). Do we need another kind of memory? *Journal of Consciousness Studies*. 27(11-12): 134-144.

Duhem, P. (1906). *The Aim and Structure of Physical Theory*. Philip P. Wiener (tr.), Princeton: Princeton University Press, 1991.

Erdelyi, M. H. (2006). The unified theory of repression. *Behavioral and Brain Sciences*, 29(5), 499-511.

Feyerabend, Paul .K., 1962, “Explanation, Reduction, and Empiricism”, in *Minnesota Studies in the Philosophy of Science III*, H. Feigl and G. Maxwell (eds.), Minneapolis: University of Minnesota Press, pp. 28–97.

Frege, G. (1892). ‘Über Sinn und Bedeutung’, in *Zeitschrift für Philosophie und philosophische Kritik*, 100: 25–50; translated as ‘On Sense and Reference’ by M. Black in Geach and Black (eds. and trans.), 1980, 56–78.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

Freud, S. (1896). *The etiology of hysteria*.

Freud, S. (1900). *The Interpretation of Dreams*.

Freud, A. (1936). Das Ich und die Abwehrmechanismen.

Grünbaum, A. (1984). *The Foundations of Psychoanalysis*. University of California Press.

Haberman, J. V. (1914). A criticism of psychoanalysis. *The Journal of Abnormal Psychology*, 9(4), 265.

Heisenberg, W. (1985). *Über den anschaulichen Inhalt der quantentheoretischen Kinematik und Mechanik* (pp. 478-504). Springer Berlin Heidelberg.

Husserl, E. (1900). *Logische Untersuchungen*

Johannsen, W. (1909). *Elemente der exakten Erblchkeitslehre*. Gustav Fischer.

Jouanna, J. (2012). The legacy of the Hippocratic treatise the nature of man: The theory of the four humours. In *Greek medicine from Hippocrates to Galen* (pp. 335-359). Brill.

Kihlstrom, J. F. (1990). The psychological unconscious. In L. A. Pervin (Ed.), *Handbook of personality: Theory and research* (pp. 445–464). The Guilford Press.

Kihlstrom, J. F. (1996). The trauma-memory argument and recovered memory therapy. In K. Pezdek & W. P. Banks (Eds.), *The recovered memory/false memory debate* (pp. 297–311). Academic Press.

Loftus, E. F. (1994). The repressed memory controversy. *American Psychologist*, 49(5), 443–445.

Lycan, W. G. (1996). *Consciousness and experience*. Mit Press.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

McNally, R. J. (2023). The Return of Repression? Evidence From Cognitive Psychology. *Topics in Cognitive Science*.

Mendel, G. (1996). *Verhandlungen des naturforschenden Vereins Brünn*.

Noreen, S., & MacLeod, M. D. (2013). It's all in the detail: intentional forgetting of autobiographical memories using the autobiographical think/no-think task. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 39(2), 375.

Otgaar, H., Wang, J., Howe, M. L., Lilienfeld, S. O., Loftus, E. F., Lynn, S. J., Merckelbach, H., & Patihis, L. (2020). Belief in unconscious repressed memory is widespread: A comment on Brewin, Li, Ntarantana, Unsworth, and McNeilis (2019). *Journal of Experimental Psychology: General*, 149(10), 1996–2000.

Poincaré, H. (1902). *Science and Hypothesis*, G. B. Halsted (tr.), Dover: Dover Publications, 1952.

Pope Jr, H. G., Schnabel, J., & Hudson, J. I. (2023). Current scientific interest in dissociative amnesia: A bibliometric analysis. *Applied Cognitive Psychology*, 37(1), 42-51.

Popper, K. (1963). *Conjectures and Refutations: The Growth of Scientific Knowledge*, London: Routledge.

Popper, K. (1974). Replies to My Critics. in Schilpp 1974: vol. 2, 961–1197.

Psillos, S. (1999). *Scientific Realism*, London: Routledge.

Quine, W.V.O. (1951). Two Dogmas of Empiricism. *Philosophical Review*, 60(1): 20–43.

Russell, B (1905). On Denoting. *Mind*, 14(4): 479–493.

Smith, S. M., & Beda, Z. (2020). Blocked and recovered memories. In *Memory Quirks* (pp. 63-82). Routledge.

Forthcoming in *Topics in Cognitive Science*. Please reference the published version of this manuscript.

Staniloiu, A., & Markowitsch, H. J. (2014). Dissociative amnesia. *The Lancet Psychiatry*, 1(3), 226-241.

Wells, F. L. (1913). On formulation in psychoanalysis. *The Journal of Abnormal Psychology*, 8(4), 217.