To remember, or not to remember? Potential impact of memory modification on narrative identity, personal agency, mental health, and well-being

Przemysław Zawadzki¹ | Agnieszka K. Adamczyk²

¹Jagiellonian University, Faculty of Philosophy, Institute of Philosophy, Krakow, Poland
²Jagiellonian University, Faculty of Philosophy, Institute of Psychology, Krakow, Poland

Correspondence
Przemysław Zawadzki, Jagiellonian University, Faculty of Philosophy, Institute of Philosophy, Krakow, Poland.
Email: przemyslaw.zawadzki@uj.edu.pl

Abstract
Memory modification technologies (MMTs)—interventions within the memory affecting its functions and contents in specific ways—raise great therapeutic hopes but also great fears. Ethicists have expressed concerns that developing and using MMTs may endanger the very fabric of who we are—our personal identity. This threat has been mainly considered in relation to two interrelated concerns: truthfulness and narrative self-constitution. In this article, we propose that although this perspective brings up important matters concerning the potential aftermaths of MMT utilization, it fails to tell the whole story. We suggest that capturing more tangible potential consequences of MMT use, namely, its psychological ramifications is crucial both in ethical considerations and in making decisions regarding the permissibility of such interventions. To this end, we first examine what current MMTs are capable of and what are the prospects of emerging MMTs. Subsequently, we outline the relationship between memory and personal identity; specifically, we indicate that concepts of self-defining memories and narrative identity are crucial to considering how MMTs may influence one's psychological functioning. On this basis, we analyze potential consequences of narrative disruption that may be the result of the use of MMTs; more precisely, we consider its potential effects on mental health, well-being, and personal agency, and outline the ethical dilemmas that decision-makers face in this context. We conclude by considering the broader cultural context that may have influence on policymaking regarding permissibility of memory modification interventions.

KEYWORDS
memory modification technologies (MMTs), mental health, narrative identity, optogenetics, personal agency, personal growth, personal identity, propranolol, well-being
INTRODUCTION

Memory modification technologies (MMTs)—interventions within the memory affecting its functions and contents in specific ways—raise great hopes but also great fears. Ethicists have expressed concerns that developing and using MMTs may endanger the very fabric of who we are—our personal identity. This threat has been framed and considered in bio- and neuroethical literature in relation to two interrelated concerns: truthfulness and narrative identity self-constitution.

As Liao and Sandberg claim, memories enable us to form a certain narrative identity, which is crucial to our having a sense of what we believe to be true about ourselves. From our past experiences and our memory of them, we may believe that we are brave or cowardly; altruistic or selfish; generous or stingy; and we may identify ourselves with certain ideologies: liberal or conservative; egalitarian or elitist; feminist or male-chauvinist, and so on. If we modify our memories in a certain way, this may change what we believe to be true about ourselves.

Some argue that this, in turn, may have radical consequences: “to deprive oneself of one’s memory—in its truthfulness also of feeling—is to deprive oneself of one’s own life and identity.” This intuition is also apparent in an influential conception of narrative identity—Schechtman’s narrative self-constitution view. According to Schechtman’s theory, having true memories must be of crucial importance, since identity self-constitution requires that a person articulate the true story about herself. Thus, since memories constitute building blocks of our stories, their modification could make our narrative identities false, and consequently make us who we are not.

This perspective brings up important matters concerning the potential aftermath of MMTs. On the other hand, it fails to tell the whole story since its primary aim is to explain the metaphysical issue of identity self-constitution. However, important, this framework is too narrow as this is not the only potential ramification of memory-modifying interventions that bio- and neuroethicists should capture. We should also understand what are more tangible potential consequences of MMT use, namely, its psychological ramifications as these are crucial both in ethical considerations and in policymaking regarding the permissibility of such interventions. Before considering these issues, it is necessary to review what current MMTs are actually capable of and what are the prospects of the emerging ones.

MEMORY-MODIFYING CAPABILITIES OF CURRENT MMTS

MMTs can modify memories through interference with the process either of consolidation, or of reconsolidation. Using MMTs to interfere with the process of consolidation means to block or alter the initial memory formation; interfering with the process of reconsolidation, on the other hand, means to intervene during a window of vulnerability in such a way that the re-storage of the memory is blocked or altered. Taking into account that it is rarely possible to administer MMTs in the aftermath of trauma, the discovery that previously consolidated (old) memories can be disrupted during a window of vulnerability occurring after their reactivation holds promise of memory-modifying interventions that may be implemented months or even years after experiencing a traumatizing event. This approach—reconsolidation-based treatment—surpasses consolidation-based interventions in several meaningful ways, mitigating the set of clinical and ethical concerns previously expressed towards MMTs.

Although it has been known for many years that reconsolidation blockade with protein synthesis inhibitors disrupts hippocampal-mediated memories—there have been means to erase even well-consolidated memories—these agents cannot be used in humans because they are highly toxic. For this reason, researchers have studied other memory-modifying drugs. Reconsolidation-based administration in humans of one of the most auspicious of them, propranolol, has been intensely investigated in recent years.

The first line of research on reconsolidation propranolol-based interventions in humans studied a clinical population of individuals...
with post-traumatic stress disorder (PTSD)—a mental disorder that may develop after experiencing trauma and for which the symptomatology may include intrusion symptoms (flashbacks, nightmares, physiological responses, uncontrollable thoughts or feelings related to a traumatic event), avoidance symptoms (avoidance of places, events, thoughts, or feelings reminding of trauma), cognition and mood symptoms, and arousal and reactivity symptoms. The results of these studies strongly suggest that propranolol might be effective in altering the reconsolidation of a reactivated trauma memory in a therapeutically relevant manner. Propranolol was shown to reduce autonomic nervous system arousal (recorded using measures such as heart rate and skin conductance response [SCR] and electromyogram [EMG] of the facial frowning muscle) to traumatic memory below the normative PTSD cut-offs as well as some PTSD symptomatology.

Another research group repeatedly demonstrated that propranolol administration in healthy populations efficiently reduces the conditioned fear response (measured by EMG as a startle potentiation) and prevents the return of the fear. However, in contrast to the studies of Brunet and colleagues, propranolol was ineffective in this research in reducing SCR. It is hypothesized that these results may reflect the possibility that startle potentiation is indicative of less conscious emotional expression that is relatively insensitive to instructions. SCR, in contrast, may depend on conscious associative learning. Thus, startle potentiation is thought to be an indicator of the conditioned fear, whereas SCR may be related to the declarative memory for the fear association. In more general terms, this may indicate that propranolol alters emotional expression of the memory, but not its content. Last but not least, this research group has also shown that promising clinical applications of reconsolidation-based propranolol administration are not restricted to PTSD. They demonstrated that it is possible to effectively diminish fear response to older and stronger memories that constitute specific phobias such as fear of spiders, snakes, and heights.

To conclude, propranolol is able to alter emotional components (or, at least, emotional expressions) of excessively emotional (traumatic, fear, or phobic) memories. Initial results suggest that a similar effect is also achievable via a non-invasive technique of state-dependent repetitive transcranial magnetic stimulation over prefrontal cortex (PFC–rTMS).

A novel MMT, optogenetics, shows an even wider range of memory modifications. To date optogenetics has been used to: activate and deactivate acquired an in vivo fearful memory, modify factual details or change the valence of memory (implant “false memory”), and to recover a “forgotten” infant memory primarily in rodents, but also in non-human primates. In animal studies, the modification of a given memory is inferred from the modification of behavior typically evoked by such memories (e.g., stress response evoked in a context in which an animal repeatedly received mild electric shocks). Thus, most of the aforementioned studies utilized a well-established fear conditioning paradigm to create a fearful memory in a studied animal whose neural underpinning could then be traced and subjected to optogenetic manipulations (for recent reviews of how this has been done see). A similar fear conditioning procedure is systematically used to study the impact of memory-modifying techniques (such as pharmacological agents) in humans.

Current optogenetic systems involve genetic modification of neurons through the insertion of opsin genes via viral infection and implantation of optical fibers with concomitant lesion of neural structures. Inserted opsins are then utilized as mediators to regulate the flow of electrically charged ions across membranes in response to light pulses. Since they are placed in specific cell types and neuronal

---

12Bisson, J. J., Cosgrove, S., Lewis, C., & Roberts, N. P. (2015). Post-traumatic stress disorder. The BMJ, 351, h6161. doi: https://doi.org/10.1136/bmj.h6161
13Although there is a discussion as to whether the findings of this research group can be attributed to interference with the reconsolidation or alteration of the memory retrieval (Elsey & Kindt, op. cit. note 9).
14Brunet, A., Saumier, D., Liu, A., Steiner, D. L., Tremblay, J., & Pitman, R. K. (2018). Reduction of PTSD symptoms with pre-activation propranolol therapy: A randomized controlled trial. American Journal of Psychiatry, 175(5), 427–432. doi: https://doi.org/10.1176/appi.ajp.2017.17050481
15Pourda, J., Sanchez, S., Tremblay, J., & Brunet, A. (2012). Trauma reactivation under the influence of propranolol: An examination of clinical predictors. European Journal of Psychotraumatology, 3, 22898386. doi: https://doi.org/10.3402/ejpt.v3i05.15470
16Brunet, A., Pourda, J., Tremblay, J., Bui, E., Thomas, E., Orr, S. P., Azzouz, A., Birnes, P., & Pitman, R. K. (2011). Trauma reactivation under the influence of propranolol decreases posttraumatic stress symptoms and disorder: 3 open-label trials. Journal of Clinical Psychopharmacology, 31(4), 547–550. https://doi.org/10.1097/JCP.0b013e3182222540; Brunet, A., Orr, S. P., Tremblay, J., Robertson, K., Nader, K., & Pitman, R. K. (2008). Effect of post-retrieval propranolol on psychophysiological responding during subsequent script-driven traumatic imagery in post-traumatic stress disorder. Journal of Psychiatric Research, 42(6), 503–506. https://doi.org/10.1016/j.jpsychres.2007.05.006
17Soeter, M., & Kindt, M. (2013). An abrupt transformation of phobic behavior after a post-retrieval amnesic agent. Biological Psychiatry, 78(12), 880–886. https://doi.org/10.1016/j.biopsych.2013.04.006; Soeter, M., & Kindt, M. (2012). Stimulation of the noradrenergic system during memory formation impairs extinction learning but not the disruption of reconsolidation. Neuropsychopharmacology: Official Publication of the American College of Neuropsychopharmacology, 37(5), 1204–1215. https://doi.org/10.1038/npp.2011.307; Soeter, M., & Kindt, M. (2011). Disrupting reconsolidation: Pharmacological and behavioral manipulations. Learning & Memory, 18(6), 357–366. https://doi.org/10.1101/lm.2148511; Soeter, M., & Kindt, M. (2010). Dissociating response systems: Erasing fear from memory. Neurobiology of Learning and Memory, 94(1), 90–91. https://doi.org/10.1016/j.nlm.2010.03.004; Kindt, M., Soeter, M., & Vervliet, B. (2009). Beyond extinction: Erasing human fear responses and preventing the return of fear. Nature Neuroscience, 12(3), 256–258. https://doi.org/10.1038/nn.2271
18Elsey & Kindt, op. cit. note 9; Soeter & Kindt (2010), op. cit. note 15.
19Elsey, J., & Kindt, M. (2017). Breaking boundaries: Optimizing reconsolidation-based interventions for strong and old memories. Learning & Memory, 24(9), 472–479. https://doi.org/10.1177/1074778917746462
20The researchers were able to achieve this by using the long-term depression (LTD) protocol. It is believed that LTD weakens synaptic efficacy and decreases spine density (Josselyn, S. A., & Tonegawa, S. (2020). Memory engrams: Recalling the past and imagining the future. Science, 367(6473), eaaw4325. https://doi.org/10.1126/science.aaw4325). Moreover, by utilizing a long-term potentiation (LTP) protocol they succeeded in retrieving the previously deactivated memory (Nabavi, S., Fox, R., Proulx, C. D., Lin, J. Y., Tsien, R. Y., & Malinow, R. (2014). Engineering a memory with LTD and LTP. Nature, 511(7509), 348–352. https://doi.org/10.1038/nature13294).
21Ryan, T. J., Roy, D. S., Pignatelli, M., Arons, A., & Tonegawa, S. (2015). Engram cells retain memory under retrograde amnesia. Science, 348(6238), 1007–1013. https://doi.org/10.1126/science.aaa5542; Ramirez, S., Liu, X., Lin, P.-A., Suh, J., Pignatelli, M., Redondo, R. L., Ryan, T. J., & Tonegawa, S. (2013). Creating a false memory in the hippocampus. Science, 341(6144), 387–391. https://doi.org/10.1126/science.1239073.
22El-Shamyaleh, Y., & Horwitz, G. D. (2019). Primate optogenetics: Progress and prognosis. Proceedings of the National Academy of Sciences, 116(52), 26195–26200. https://doi.org/10.1073/pnas.1902984116.
23Adamczyk, A. K., & Zawadzki, P. (2020). The memory-modifying potential of optogenetics and the need for neuroethics. NanoEthics, 14, 207–225. https://doi.org/10.1007/s11569-020-00377-1; Josselyn & Tonegawa, op. cit. note 20.
24Elsey & Kindt, op. cit. note 9.
pathways, their expression enables optogenetics to activate or reversibly silence (inhibit) specific neural circuits. Because of the therapeutic potential of optogenetics (which is not restricted to memory modifications), there is an intensive effort to design optogenetic systems that are sufficiently safe and non-invasive to be used in humans. The progress of this research is rapid. Thus, although optogenetics’ transformation from an investigatory tool utilized in animal models into therapeutic MMT in humans will take at least a dozen years, it is important to start considering whether there may be inherently undesirable neuroethical consequences of optogenetics, among other memory-modifying technologies.

3 | PSYCHOLOGICAL CONSEQUENCES OF NARRATIVE IDENTITY DISRUPTION VIA MMTS

Singer et al.’s discussion of narrative identity is useful to consider how memory modifications may influence a person’s psychology. Singer et al. argue that self-defining memories—autobiographical memories constitutive of the most important and recursive themes in a person’s life—and narrative scripts—a person’s rules for predicting, interpreting, and responding to a set of scenes (i.e., events, memories)—serve as the ingredients for the overall life story of a person. These components, in turn, make up the narrative identity of a given individual. To further elucidate their model, they propose to think of the pool of self-defining memories and scripts as possessing a synchronic existence within narrative identity; they accumulate associations and connections, while remaining relatively static in the personality. One can think of the life story as diachronic—it moves forward in time and continually amends itself in light of new experiences.

One of the most general questions from a neuroethical perspective that may arise in this context is: why should we consider a narrative valuable, or fear its possible disruptions by memory modifications? As we noted in the Introduction, influential accounts have argued that a narrative is a constitutive condition of the self. In the next subsections, we provide different answers to this question by analyzing more tangible consequences of the narrative disruption that may be the result of MMT use.

3.1 | Transformational power of self-defining memories and their contribution to personal growth and happiness

In recent years, psychologists have explored the relationship between narrative identity and adaptation. Investigations in this area show that individuals who find redemptive meanings in suffering and reversals and who construct narratives emphasizing personal agency and self-exploration usually enjoy better mental health and overall well-being. Many instances of redemption sequences are present in the life stories of middle-aged adults who score high on self-reported measures of generativity (indicator of a strong commitment to improving society and promoting the well-being of future generations). In redemption sequences, demonstrably “bad” or emotionally negative events lead to demonstrably “good” or emotionally positive outcomes. The initial negative state is thus “redeemed” by the greater good perceived by the self as its ultimate result. Redemption indicates the transition in a life narrative from an emotionally negative event to a positive outcome or positive attribution in relation to the self. The adaptive function of conceptualizing lives as stories of redemption may be reliant on sustaining the hope or confidence needed in trials and tribulations while reinforcing longer-term commitments.

Therefore, if an individual were to decide to remove (with the use of optogenetics) or to change or dampen the emotional component (with the use of optogenetics, propranolol, or PFC–tRMS) of negative memories, she might deprive herself of the opportunity to experience a redemption sequence. Thus, such interventions might deprive an...
individual of the potential for finding meaning in suffering and prevent her from reshaping emotionally negative events into positive outcomes such as self-transformation, or involvement in prosocial goals, for example, helping others. Therefore, the concept of redemption sequences suggests that to make someone’s life more fulfilling, it may not be sufficient merely to influence memory through MMT. It may be that we often need to come to terms with more difficult times in our lives to transform ourselves—an idea consistent with a long humanistic tradition emphasizing the character-shaping role of hardships.

The redemption sequence points to the broader adaptive problem of how humans make sense of suffering in their lives in narrative terms. Research on narrative identity suggests that adults who emerge stronger or enhanced due to negative life experiences tend to engage in a two-step process: the first associated with personal growth, the second with happiness. In the first step, the negative event is re-experienced and reconsidered. The individual explores questions concerning how the event came about, what it might have led to, and what role it might play in her autobiographical story. In the second step, the individual articulates and commits herself to a positive resolution of the experience.

Thus, a person who undergoes erasure or change of the valence of the negative (or even traumatic) memory might miss an opportunity to undertake self-exploratory measures concerning her experiences. She may, therefore, be unable to take advantage of bad experiences by learning a life lesson and gaining valuable insight into a deeper layer of herself. In more naturalistic terms, the individual may miss the opportunity to strengthen her “psychological immune system” and thus, deprive herself of the positive effects of overcoming negative experiences, as many studies demonstrate that positive resolution of negative events is associated with higher levels of happiness and well-being.

Traumatic experiences often constitute self-defining memories. Self-defining memories are postulated to be “affectionately charged reconstructions of past events that have the power to shake our rational understanding of past experiences, [and] bias our ongoing processing of information.” Thus, erasure or “pacification” of self-defining memories may infringe these powerful lenses, a person’s interpretations, and with them her experiences, interactions, values, and reactions could become altered. This may endanger a person’s authenticity—the very sense of being oneself (for works devoted exclusively to this issue see).

Modifications of self-defining memories may also lead to encroaching a person’s narrative scripts. In psychotherapy, overcoming adverse narrative scripts can have positive effects for an individual—it can allow him to introduce “new behavior and lead to different endings to situations that would otherwise have repeated his familiar pattern of disappointment and humiliation.” Modifying narrative scripts by altering memories that constitute them with the use of MMTs, may, however, undermine efforts of the individual's work on self-reflective overcoming of a negative experience and not let new behavioral scripts develop during this process.

Assume that a person gives unconditional support and trust to his husband, and, as a result, he experienced exploitation and betrayal. A painful memory that is a result of this betrayal may serve as a lever to overcome narrative scripts of this “unconditionality.” More specifically, one may decide to stay with one’s partner, but only on the condition that he will never cheat on him again, and if he decides to break that condition, their relationship shall come to an end. So, if he was to erase the memory of his partner's past betrayal, he risks repeating the same pattern of behavior as he does not learn from the experience, thereby precluding change of his behavioral scripts—“those who do not learn from history are doomed to repeat it.”

On the other hand, Liao and Sandberg argue that even if we remove a declarative memory of an event, the procedural memory will still remain. That procedural memory could be sufficient to protect a person from committing the same mistake again, as one would follow the “intuition,” “hunch,” or “somatic marker” in the neuroscientific nomenclature that develops in addition to factual memory of an event. Returning to our example, imagine that a betrayed husband forgives his partner on the condition that he will never betray him again, but this time he also decides to remove a declarative memory of the whole affair to “clean the slate.” Applying the logic of Liao and Sandberg, were the unfaithful partner to cheat on him in the future, procedural memory (“hunch” or “gut feeling”) would protect the betrayed husband from repeating the same mistake of trusting and forgiving. So, in the case of subsequent betrayal (s), the exploited partner would decide not to remove his memory, but instead leave his unfaithful husband.

However, it might be questionable if one would choose to follow his hunches and gut feelings in making such an important decision as leaving one’s partner, as one could be committed to being a rational person who does not base his decisions on gut feelings or “sixth
Remember in this context appraisal theories of emotion and the classic "bridge" experiment on misattribution of arousal (Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69(2), 379–399. https://doi.org/10.1037/h0046234; https://en.wikipedia.org/wiki/Misattribution_of_arousal).

Moreover, it seems plausible that lacking the knowledge of one's partners history of betrayals, one could easily misattribute his "hunch" feelings and interpret them as signs of deep suffering that have resulted from a current betrayal, rather than an indication that he has been deceived for the second time. Finally, when trying to find an explanation for the partner's ambiguous behavior (for instance, his late arrival), one might undermine potential "hunch" feelings as irrational and paranoid, and thus be motivated to suppress them. Consequently, when faced with the second betrayal, the betrayed husband might either decide to (again) come to an agreement, forgive his partner and decide to (again) remove his memory of this unfortunate incident, or to completely ignore all the "warning signs" that could be an indication of infidelity.

Let us now consider an alternative scenario of removing only the emotional component of the betrayed husband's memory, as were he to decide to give another chance to this relationship, he might be interested in relieving himself of the emotional burden associated with his partner's betrayal, rather than removing the declarative memory of that event. In this scenario, we argue, breaking the spell of dysfunctional behavioral scripts of a betrayed husband is also unlikely. Cognitive (neuro)science undermines a long-held view of people as predominantly rational agents and tends to emphasize the role of emotions in both learning and decision making. In theory, learning from experience could be based solely on the factual content of bad memories, but in practice, negative affective content seems indispensable to prevent following the same dysfunctional pattern of behavior. Returning to our example: knowledge that a husband committed infidelity may not be sufficient to leave him, as a strong emotional reaction—a deep pain—is required for a person to make such a difficult and important life-changing decision.

To conclude, when emotions are divorced from knowledge (as could be the case when one would remove declarative content of memory), they appear to be degraded in intensity to mild, free-floating "hunches" or "gut feelings" that can be easily misinterpreted, misattributed, or undermined as being unsupported by relevant "evidence." So, although hunch feelings might be enough to push one in the right direction when it comes to decisions that are, to a large extent, reached automatically (such as whether to put the pedal to the metal or to drive responsibly), it may be that strong, fully-fledged emotions (that have an unambiguous and easily identifiable source) are needed to reach important, life-changing decisions. On the other hand, when knowledge is divorced from emotions (as could be the case when one would remove an affective component of memory), it may also lack the motivational power to influence one's decisions, as people appear to depend heavily on their built-in emotional compasses in addition to or independently from the possessed knowledge. Finally, it appears that both ingredients are necessary to bring a substantial change into one's life narrative as revealed by the transformational power of highly emotional experiences. Therefore, the prospect of erasure of only the declarative or only the emotional component of a memory does not seem to prevent a person from repeating the same mistakes again and may actually be an obstacle for the development of new, healthy behavioral scripts to deal with adverse situations in one's life.

In summary, although blunting the emotional impact of one's memories or erasing them altogether might seem a tempting way to relieve current suffering, it seems to bring only a short-term resolution, ultimately blocking the opportunities for personal growth and leading to greater long-term unhappiness as one fails to introduce necessary changes to one's life and instead preserves the status quo. It is important to note that failing to adjust one's behavior to new circumstances often not only exacerbates the existing problems, but also might lead to new ones, creating a negative feedback loop. For instance, in the case of the betrayed husband, his forgiving attitude and unconditional love after repeated removal of his memories could lead to disregard and more frequent and blatant betrayals by his partner, who could eventually decide to dump such a naïve person, for whom he has lost all respect—a prospect that would arguably bring more pain than initiating the end of the relationship on one's own. Conceivably, this could also affect relations with his friends and family, who could, for instance, start avoiding him, also losing respect for someone seemingly so desperate to stay in such a dysfunctional relationship, or, simply not knowing how to behave in these unusual circumstances (e.g., whether to break the news to him or remain silent and watch how he is being taken advantage of by his partner). It seems that, in cases of modification involving the erasure or deactivation of the valence of adverse memories, we are dealing with a difficult calculation between the pros and cons of overcoming painful memories on one's own (or with the help of a psychotherapist) versus deleting or modifying the source of suffering, that is, memory. On one hand, it seems obvious that clinicians should respect a patient's autonomy and allow one to make an independent decision regarding modification versus an attempt to overcome a given memory on her own. On the other hand, patients may too often opt for memory modification on the basis of their current mental suffering. It is a well-established psychological truism that most people tend to put the interests of the current self before those of the future self. Furthermore, this preference seems all the more likely in the context of memory modifications given that research shows that establishing the meaning of personal narratives is hard work and may prove emotionally costly.

Therefore, the difficult question arises as to whether clinicians should take a paternalistic approach in situations of this kind. The

---

42 Remember in this context appraisal theories of emotion and the classic "bridge" experiment on misattribution of arousal (Schachter, S., & Singer, J. (1962). Cognitive, social, and physiological determinants of emotional state. Psychological Review, 69(2), 379–399. https://doi.org/10.1037/h0046234; https://en.wikipedia.org/wiki/Misattribution_of_arousal).

43 Kaheman, D. (2011). Thinking, fast and slow. Farrar, Straus & Giroux; Ariely, D. (2008). Predictably irrational: The hidden forces that shape our decisions. HarperCollins Publishers.

44 An example of Liao & Sandberg, op. cit. note 1.

45 Damasio, op. cit. note 40.

46McAdams & Jones, op. cit. note 35; Wilson, J. P. (2007). The posttraumatic self: Restoring meaning and wholeness to personality. Routledge.

47 Many thanks to Andrea Lavazza, whose insights helped us to develop this argument.

48McAdams, op. cit. note 29.
answer is all the more difficult because no one can know a priori whether a given memory is pathological, or whether a person can cope with it by incorporating it into the adaptive narrative, thus enriching herself in the long term. It appears that the issue of what is adaptive for a given person as opposed to what is purely harmful depends on the individual characteristics of a person. Therefore, the potential for a strict demarcation between adaptive and pathological memories that can be applied to the entire clinical population appears extremely difficult, if not illusory. Thus, it seems that the decision of whether to intervene should be individualized/adjusted to a particular case. More specifically, what should, at the very least, be taken into account are one's system of values, life goals, and long-term aspirations whose achievement might have been hindered or undermined by a traumatic (memory of an) event (the issues on which we elaborate more in the next subsection). This is because—as we mentioned earlier—traumatic events have a very strong transformational power and they often become self-defining memories. In a broader perspective, such decisions would probably also reflect, to some extent, the values that would dominate in society, when MMTs become a more widespread form of intervention. For instance, in the above discussion, the relevant values to be discussed are along the line of autonomy (motivated by the care about the freedom of choice of the person's current self)/paternalism (motivated by the care about the person's future self). We return to this topic and discuss it from a broader perspective in the Conclusions.

3.2 Personal agency, mental health, and well-being

Altering personal agency is another crucial issue in the context of potential narrative disruption that may be the effect of modifying painful memories. Removal or emotional modification of difficult memories may result in diminishing a patient’s sense of agency, that is, her feelings and beliefs about her ability to control herself, the world, and others, and to make her own decisions. We propose to distinguish between the normatively and psychotherapeutically relevant components of this concern.

The former refers to the idea that an individual should respect herself as an agent. Agency can be either viewed as an autotelic value or explicated in more naturalistic and functional terms. In the former perspective, it is important to believe that we are creatures with free will and the ability to determine our own destiny. Thus, it can be argued that by removing a memory or deactivating its negative emotional component by means of MMTs, the individual misses an opportunity to think through the experience in question and to ultimately overcome it on her own terms. Thus, she gives up self-determined agency for the sake of more hedonistic values such as “peace of mind.” On the other hand, functionalists may refer to the adaptive dimension of agency. In simplified terms, a more naturalistically inclined ethicist may argue that every time one takes an action one learns something new; when one fights hardships (instead of giving up), one's fitness increases over the long term. Thus, agency is valuable from the adaptive perspective.

The psychotherapeutic aspect of the issue of personal agency in the context of memory modifications refers to the relationship between the process of incorporating memories with negative emotional valence into the autobiographical story and mental health. In Adler's studies of psychotherapy treatment, enhancement in patients’ perception of their own personal agency preceded and predicted alleviation of symptoms and improvement in mental health. As patients constructed stories that increasingly emphasized their ability to control their surroundings and make their own choices, they showed corresponding reductions in symptoms and improvements in mental health. Therefore, the opportunity to think through the memory of an unpleasant event appears important not only because it incorporates respect for oneself as an agent, but also because, if executed successfully (i.e., if incorporated into the adaptive narrative story emphasizing one’s agency in overcoming hardships), it may improve the patient’s mental health.

On the other hand, it appears that the only possible solution to overcoming a traumatic memory is meaning-making, which materializes in committing oneself to turn a negative experience into something positive, for example, by helping other victims of similar trauma/abuse or raising society’s awareness of a problem one might have experienced themselves (consider the #MeToo movement or Sharon Tate's mother's efforts). Consequently, engaging in the process of overcoming the trauma on one's own (i.e., without the use of MMT) may paradoxically narrow down one's options, enforcing the re-evaluation of one's goals in life, and ultimately diminish one's agency. Moreover, some experiences may be so painful that an individual could be left incapable of not only constructing a satisfying narrative, but even leading a functional life due to the adverse memory. Some claim that recurring traumatic memories resulting in the development of PTSD constitute a paradigmatic example of hijacking of narrative identity. In such cases, personal agency and autonomy may be augmented by MMTs by offering a "fresh start." This argument applies primarily to clinical cases. However, let us consider two cases in which the use of MMTs could also be justified by non-clinical objectives—the first deals with not wanting to give up one's previous life and the second one with not being able to embrace one's new life.

Consider a person, Nietzschean, who does not want to pursue prosocial goals, such as helping others, and might instead want to pursue more egoistic goals, such as wealth, fame, and power. Imagine that Nietzschean experiences some traumatic event. In such a

50 Adler, J. M. (2012). Living into the story: Agency and coherence in a longitudinal study of narrative identity development and mental health over the course of psychotherapy. Journal of Personality and Social Psychology, 102(2), 367–389. https://doi.org/10.1037/a0025289
51 McAdams, D. P. (2019). Psychopathology and the self: Human actors, agents, and authors. Journal of Personality, 88(1), 144–155. https://doi.org/10.1111/jopy.12496; Hobfoll, S. E., Gaffey, A. E., & Wagner, L. M. (2020). PTSD and the influence of context: The self as a social mirror. Journal of Personality, 88(1), 76–87. https://doi.org/10.1111/jopy.12439
52 Liao & Sandberg, op. cit. note 1.
scenario, to successfully deal with her traumatic (or very negative, yet not clinically relevant) experience, she would be left with no choice but to create an effective redemption story, which would enforce the incorporation of her trauma into her narrative self. However, in the case of Nietzscheana, it is very likely that this would necessitate a change in her previous life goals, system of values, and aspirations. Eventually, Nietzscheana would end up with fewer potential life stories that she could create than before her trauma. Although one could argue that it would actually be a “good” change—a kind of moral development that should be encouraged as it would likely have positive consequences for society—from the perspective of Nietzscheana, it could be best to remove a traumatic memory to get her “back on track” to achieve goals she was striving for before her traumatic experience.

The above considerations illustrate a case when one might not want to give up one’s previous life (goals) after an unexpected and potentially life-changing event, and may object to the idea of transforming oneself to successfully cope with a trauma. However, there may also be cases when one has successfully dealt with a trauma in terms of (not developing) clinical symptoms, but nonetheless, past memories are hindering full embracement of one’s self-determined life-path.

Remember in this context the case of J. D. Vance, the protagonist of Hillbilly elegy (2020). It might be true that had he not been exposed to negative experiences during his childhood, he would not have had such strong motivation to become a top Ivy League student. Imagine, however, that at some point in his life, when J. D. is a prosperous adult, he finds himself haunted by the most painful memories of his troubled past (e.g., remember the scene when his mother tried to kill them in a car accident). He feels alienated from his middle-class friends, who do not share his experiences of poverty, addiction, and abuse (remember the scene at a dinner table when J. D. was trying to secure his internship). Eventually, he starts to suffer from impostor syndrome and feels as if he “doesn’t belong” to his middle-class life that he pursued so relentlessly. Although erasing all memory of his childhood would obviously be an ill-advised idea, J. D. might wish to remove some of his most negative memories, which could enable him to shake off his feeling of “inadequateness” and estrangement. Arguably, doing so could enhance his agency and autonomy in shaping his preferred narrative, and likely contribute to his well-being and psychological health (remember results of Adler’s study as discussed above).

Both of the above situations are strong cases for arguing that MMTs should also be permissible in non-clinical cases. However, it is possible to offer, in the context of personal agency, a radical and general argument against memory modifications. This would apply not only to negative memories, but to all kinds of autobiographical memories. The argument is this: once an individual deprives herself of given memories (or even their emotional complexion), she gives up the opportunity to autonomously decide what she wants to incorporate into her autobiographical narrative and overall understanding of who she is in the future. In other words, memory modifications may constitute points of no return as such interventions inescapably deprive a person of “raw materials” (i.e., memories) from which she could construct her narrative identity. Obviously, this is on the assumption that learning about one’s past from others’ testimonies does not have the same epistemic value as the first-personal experience and memories.

However, the above-discussed radical argument against the use of MMTs seems inappropriate in the context of forgetting and repressing traumatic memories—people lose content or access to memories and this is the natural course of things, so we should not require the person we consider autonomous to always have all the possible material from which to build the narrative. Moreover, specific characteristics of the above discussed MMTs—that is, the capacity of optogenetics demonstrated by Nabavi et al.52 to repeatedly deactivate and reactivate a specific memory by modifying synaptic strengths as well as evidence that emotional aspects of adverse memories can be relearned after reconsolidation-based administration of propranolol53—make the above-mentioned radical proposition lose its grievousness. Since the effects of both optogenetics and propranolol could arguably be reversed, it precludes the problem of the irreversible infringement of personal agency. Finally, it seems that users of MMTs could even derive at least a certain sense of agency from the ability to shape their memories (and, by extension, their affective life) according to their will.55

In sum, it should be pointed out that the decision to use MMTs should necessarily take into account the severity of symptoms and suffering associated with recalling a given negative event and the chance that the individual can handle them. If the patient feels (or the doctor knows based on the most complete scientific evidence) that she will not be able to free herself from a given traumatic memory by conventional measures, not only mental health considerations (e.g., risk associated with the chance of developing PTSD), but also well-being and personal agency considerations may be sufficient reasons for opting for MMTs. As we have pointed out, erasure or emotional modification of debilitating memories may actually be the only way to give a subject control over her life and story, thereby allowing for a revival of meaningful narrative identity and enhancing her sense of agency.

4 | CONCLUSIONS

We conclude with more universal considerations regarding the emerging potential of MMTs to modify memory. As shown by, for example, the mechanism of repression or forgetting, selective

---

52 For an argument, see Liao & Sandberg, op. cit. note 1.
53 Nabavi et al., op. cit. note 20.
54 Soeter & Kindt (2011), op. cit. note 15.
55 Many thanks to the anonymous reviewer who suggested this argument to us. Alternatively, one can argue that by deciding to tinker with her memories, a person could feel “defeated” by her own psyche, unable to cope with her problems on her own and having to resort to some external intervention, which could diminish her sense of agency.
Rewriting of human lives appears to be the natural propensity of our cognitive systems, wired into our brains by evolution. Therefore, by refining memories through neuropharmacology (e.g., propranolol, glucocorticoids) or neurotechnology (e.g., optogenetics, PFC–rTMS), we could acquire the means to do this more accurately and effectively, providing hope to those less endowed with these kinds of abilities by nature, and to those for whom this is impossible due to neurological or psychiatric conditions. On the other hand, we should be cautious when it comes to tinkering with memories, since—as discussed above—this may have consequences for the identity, agency, well-being, and mental health of the person in question.

Deliberate and selective modifications of painful autobiographical memories become more and more possible in a clinical setting. However, the decision to intervene cannot always be based on purely scientific considerations, as these interventions involve several important normative aspects, such as questions about personal identity, authenticity, or autonomy. Moreover, there is still no scientific consensus concerning the manner in which traumatic memories should be treated as the consequences of memory modifications are still largely unknown. Furthermore, as is apparently always the case with controversial medical interventions, decisions to interfere with a person’s memory will be determined to some extent not only by explicit ethical and scientific considerations, but also by prevalent ideologies, such as scientific humanitarism or transhumanism, which will most often be in favor of a cure, and more traditional humanistic and religious values, which will usually be against it. Therefore, if it becomes possible to modify an individual’s autobiographical memory “at will,” then the permissibility of this will depend on a large extent on values of a dominating ideology, and be a result of subtle social pressures and the zeitgeist. For example, given the more and more capitalist mode of being towards which our culture is heading, and having in mind how deeply capitalism has already colonized the thinking of the modern human, MMTs may open up a rather dystopian possibility to adjust the most intimate contents of our minds (memories) to fulfill certain goals (e.g., successful careers) and norms (e.g., efficiency) via technological means for appropriate fees. On the other hand, the concern for the truthfulness widely discussed in neuroethological literature (see Introduction) may serve as a counteracting force. Although, one could worry, since we live in the era of post-truth politics—at a time when the truth is not seen as an absolute value but merely an instrumental one—that the truth may not matter too much in this process.

Taking all of this into consideration, in designing future policies, we should not only review the existing scientific knowledge regarding the potential side effects of MMTs, but also bear in mind that decisions to intervene in an individual’s memory cannot (and will not) be based purely on scientific considerations. Thus, the role of bio- and neuroethics should be also to reveal the impact of prevailing ideologies in influencing these decisions, analyze which values are at odds while making them, and invite a broader audience to well-informed discussion regarding the policy of memory modifications, as this debate will decide whether and in what circumstances MMT use will be allowed.

Acknowledgements

Open access license of the publication was funded by the Priority Research Area Society of the Future under the program “Excellence Initiative–Research University” at the Jagiellonian University in Krakow.

Conflict of Interest

The authors declare no conflict of interest.

Author Biographies

Przemysław Zawadzki majored in Philosophy and Cognitive Science. He is a PhD candidate in the Philosophy and Jagiellonian Interdisciplinary PhD Programme at Jagiellonian University in Krakow. Currently, he is working on his research grant “Self in the Context of Deep Brain Stimulation (DBS) Systems. Issues of Personal Identity, Authenticity, Autonomy and Responsibility” under the “Diamond Grant” Programme. He is primarily interested in neuroethics, philosophy of mind, and the evolution of morality.

Agnieszka K. Adamczyk is a graduate of Psychology and a PhD candidate in the International PhD Programme in Social and Cognitive Neuroscience (CogNeS) at Jagiellonian University (PL) and Radboud University (NL). In her doctoral thesis, she studies emotions with the use of neurophysiological and neuroimaging methods (such as EEG, fMRI).

How to cite this article: Zawadzki, P., & Adamczyk, A. K. (2021). To remember, or not to remember? Potential impact of memory modification on narrative identity, personal agency, mental health, and well-being. Bioethics, 35, 891–899. https://doi.org/10.1111/bioe.12926

56 Lavazza (2019), op. cit. note 3.
57 Or compulsoriness, in certain circumstances in societies where autonomy is not highly valued.