SAPIENZA
UNIVERSITÀ DI ROMA

E-ISSN 2531-7288
ISSN 0394/9001

MEDICINA NEI SECOLI
Journal of History of Medicine and Medical Humanities
34/2 (2022) 103-128

Received: 19.11.2021
Accepted: 01.02.2022
DOI: 10.13133/2531-7288/2652

Corresponding author:
Alexandre Erler
alexandre.erler@philosophy.oxon.org

Memory Erasure and the Objection from Truthfulness

Alexandre Erler
National Yang Ming Chiao Tung University, Taipei, Taiwan

ABSTRACT

Memory Erasure and the Truthfulness Objection

The prospect of selectively erasing undesired memories, whether inducing trauma or "normal" negative affect, has long been explored in fiction. Today, advances in biomedical science increasingly promise to turn it into reality. This article discusses one particular ethical concern about memory erasure, premised on the value of "truthful living". After explaining memory erasure (alongside other forms of memory editing) and reviewing its current science, I lay out what I call the truthfulness objection. I then consider two main challenges to it: a skeptical take on the accuracy of autobiographical memories (which I critique), and a challenge to the normative force of truthfulness (which I partly endorse). After highlighting what I take to be the grain of truth in the objection, I conclude on a cautiously optimistic note, by highlighting some practical constraints that can be expected to reduce the threat to truthfulness from memory erasure.

Keywords: Autobiographical memory - Memory erasure - Neuroethics - Truth

1. Introduction: the ethical debate around memory editing

"Memory editing" broadly refers to the process of altering a person's memories (other than by *enhancing* memory capacity) using direct interventions into the brain, with a view to improving her well-being. Prospective interventions that would allow to either selectively erase painful memories, or to blunt their emotional impact, have generated a substantial literature in neuroethics over the past two decades. These interventions are primarily

being considered for therapeutic purposes: for instance, for the alleviation of conditions like Post-Traumatic Stress Disorder (PTSD) or phobias. However, the possibility of using them to simply alleviate the "normal" pain from non-traumatic memories has also received considerable attention[1].

The present paper discusses one particular kind of memory editing, currently still at the stage of research and development: memory erasure (of the *voluntary* kind, not coerced or unintentional), especially using biomedical interventions. Procedures like memory erasure raise a range of ethical concerns. These include possible threats to personal identity, or at least to the coherence of a person's self-narrative, through the creation of discontinuities in her mental life[2]; the risk of corrupting a person's emotional responses to her past[3]; and interference with a person's capacity for autonomous living via disruptions to her values[4], among other issues. In this paper, I will focus on a particular objection to memory erasure that appeals to the value of truthfulness or "truthful living". After outlining the various possible forms of memory editing, and highlighting relevant distinctions between different types of memory, I will present an overview of the current science of memory erasure (and associated interventions, such as false memory implantation). I will then lay out what I call the truthfulness objection to memory erasure, which broadly states that the value of truthful living speaks against erasing undesired memories - and *decisively* so in cases where the procedure does not alleviate a pathological condition. I will consider two main challenges to that objection: first, a challenge to the idea that erasing memories would make our lives less truthful, based on a skeptical view of the accuracy of autobiographical memories; and secondly, a challenge to the normative force of truthfulness, arguing that it can be outweighed by competing considerations, and that memory erasure can be permissible, even when used non-therapeutically.

I will contend that while the skeptical challenge makes some valid points about the limits of the reliability of autobiographical memory, it nevertheless goes too far in questioning that reliability, and minimizing the foreseeable impact of full memory erasure on truthful living. As for the critique of the normative force of truthfulness, I will largely agree with it, while nevertheless arguing that the truthfulness objection does have the merit of highlighting the often *suboptimal*, albeit not *ethically impermissible*, nature of memory erasure, especially for non-therapeutic use. I will conclude on a cautiously optimistic note, by highlighting some practical (and especially social) constraints that can be expected to reduce, while not completely removing, the threat to truthfulness presented by memory erasure.

2. Memory erasure and other forms of memory editing

Major examples of memory editing include:

1. Memory erasure;
2. Memory blunting;

3. The implantation of false memories;
4. Switching the affective valence of a memory: for instance, causing a particular memory to elicit pleasant rather than unpleasant emotions.

Categories 1 and 3 are of particular relevance to the issue of truthfulness, on which the present discussion will focus. Importantly, however, the distinction between memory erasure and memory blunting should not be viewed as too clear-cut. The general idea behind it is that we can contrast interventions that either erase all or part of the content of a memory, or series of memories ("memory erasure"), with those that only reduce a memory's emotional impact on the person, while leaving its factual content untouched ("memory blunting"). A person who blunted, rather than erased, her memory of a traumatic event, would thus still remember the details of what she witnessed just as accurately, yet her recollection would no longer "sting" as much. For example, in the kind of scenario that much of the scientific research on memory editing is seeking to achieve, her memory would no longer cause her to experience the symptoms of PTSD. Yet a complication here is that one way of achieving memory blunting involves the erasure of memories of a certain kind. To clarify this, we need to introduce some further distinctions, standard within the psychology of memory.

The first such distinction is between *declarative* and *non-declarative*, or *implicit* memory. Declarative memory allows for the conscious recollection of events and facts[5]. By contrast, implicit memory refers to various forms of learning that, unlike declarative memory, operate outside of conscious awareness and can result in the formation of different skills, habits, and dispositions[6]. These might include, for instance, knowing how to ride a bike, but also certain learned emotional dispositions, such as the conditioned fear responses induced in many of the studies that form the science of memory editing. Within declarative memory, a further distinction is drawn between *episodic* and *semantic* memory. Whereas episodic memory is defined as "memory for specific experiences, usually associated with a time, place, and emotion" semantic memory concerns "the recollection of facts and generalized knowledge about the world"[7] - although it can also include knowledge about oneself[8].

Paradigm cases of "memory erasure", those that the phrase tends to bring to most people's minds, will involve the erasure of a person's painful *episodic* memory (and usually, also the associated semantic memories). One famous example in fiction is the 2004 movie *Eternal Sunshine of the Spotless Mind*, in which an estranged couple, Clementine and Joel, both resort to the services of a firm named Lacuna Inc. to erase their memories of each other, after their relationship has turned sour. However, at least some forms of memory blunting will also involve erasing a certain type of memory, namely an implicit memory (such as a fear memory), while seeking to retain the declarative memories (episodic and semantic) of the event that gave rise to it[9].

In principle, memory erasure could be combined with the implantation of false memories, in order to prevent it from leaving any confusing gaps inside a person's inner

story. Yet false memories could also be introduced in the absence of memory erasure. There are already familiar psychological techniques for planting false memories into people's minds, such as simple exposure – which could also happen accidentally - to misleading information[10]: say, causing the witness of a robbery to falsely remember having seen a blue car by subsequently asking them "how many suspects got out of the blue car", when the car they saw was in fact white. The kind of misleading information that might be used to generate false memories is becoming increasingly diverse, now also including doctored photographs[11] and, most recently, "deepfake" videos[12].

While the use of such methods by malicious actors to implant false memories without a person's consent is clearly a worrying prospect, their hypothetical *consensual* use by people to improve their own well-being raises questions about feasibility. Some have thus suggested that the fostering of false memories about food or alcohol consumption (e.g. an inaccurate memory of a very unpleasant experience drinking alcohol) might help some people adopt healthier lifestyles[13]. However, concerns might arise if it turned out that medical professionals could not successfully implement such a proposal while securing their patients' informed consent beforehand. The question would then be whether this dilemma could be resolved by having prospective patients consent to some form of deception, without necessarily being told about the specifics of the method to be employed, so as to avoid undermining the process of memory implantation[14]. Further empirical evidence would be needed to answer that question. In addition, individual differences exist in the proclivity to develop false memories[15], meaning that such traditional methods of false memory implantation might fail to work in many cases.

In the future, it is conceivable that such technical limitations will be overcome by using technological means to implant false memories, somewhat like in the 1990 sci-fi movie *Total Recall*. MIT neuroscientist Steve Ramirez and his colleagues may have taken a step in that direction when, using the cutting-edge technology of optogenetics (which involves the use of light signals to control the activity of specific neurons), they implanted a false fear memory into a mouse. The mouse was thus made to falsely remember having received electric shocks in a context in which it had never been shocked, as a result of optogenetic activation of relevant neurons in the different context in which it actually received the shocks[16]. That being said, it is still a very big step from such studies to the implantation of false memories of the *Total Recall* kind. For one thing, it is significantly more challenging to apply techniques like optogenetics in humans than in mice, partly due to the risks presented by the use of such an invasive procedure[17]. For another thing, implanting a false *implicit* memory, such as the false fear memory that Ramirez's team induced in mice, is clearly different from implanting a false *episodic* memory, as portrayed in *Total Recall*. Episodic memories include, for instance, rich forms of mental imagery that are absent from implicit memories. Nonetheless, since conventional psychological techniques are already able to generate

false episodic memories, there is reason to believe that such a feat need not always remain beyond the reach of higher-tech tools like optogenetics.

The primary cases for which memory editing is being considered are of the therapeutic kind. For instance, people suffering from PTSD, such as members of the armed forces returning from combat, would be prime candidates. There is also hope that memory editing can help combat anxiety disorders and addictions. However, as illustrated by movies like *Eternal Sunshine*, it seems that quite a few people might be interested in editing unpleasant memories to improve their subjective well-being, even when these do not give rise to a psychological disorder. This might involve removing one's memories of a past, failed relationship, or of highly embarrassing or humiliating past acts or experiences.

Given how radical a procedure the full erasure of an autobiographical memory, or set of memories, seems to be, some might wonder why one would ever consider taking such a step, if the more moderate option of blunting the emotional impact of the relevant memories is also available. A first reason has to do with the potential limitations of current methods of memory blunting. For instance, there is ongoing debate as to whether existing approaches that seek to blunt the "sting" of long-standing memories by targeting the "reconsolidation" process[18], actually manage to erase the implicit memories that they are targeting, or rather simply "bury them deep" without removing them. If the latter happens to be true, this would raise a concern about the possibility that the targeted memories might resurface in certain circumstances, leading to a return of the undesirable emotional responses. This concern already applies to behavioral methods of memory blunting like extinction training[19].

Secondly, it can be argued that even successfully removing an implicit memory associated with an unpleasant past experience cannot be expected to *completely* alleviate the negative affect that the recollection of that experience might elicit. Even if – as one might hope! – such a procedure could successfully extinguish traumatic or phobic responses, it still seems plausible to think that the factual content of declarative memories, in interaction with the person's moral compass, will at least often be enough to trigger unpleasant emotional responses in the person concerned. Such responses may not result in a full-fledged mental pathology like PTSD; yet the knowledge, for instance, that one had previously done something particularly reprehensible or embarrassing (especially, perhaps, if one clearly remembered the details of committing that act), would alone surely cause many to experience intense guilt and embarrassment. Short of somehow blunting - say, through the use of antidepressants - the general propensity to experience these emotions across the board, avoiding those unpleasant responses might require erasing the declarative memories of the event.

It might be objected here that there is at least as much cause for skepticism about the feasibility of memory erasure as comprehensive as that undergone by the protagonists of *Eternal Sunshine*. Thoroughly wiping out a former romantic partner from one's

mind would thus require removing not just one memory but many, which in turn would be deeply entangled in a web of further memories, such as the memories of the various events one experienced while in the company of that person[20]. This is indeed an important point. That said, removal of a single, particularly significant declarative memory, or of a limited set of such memories, looks like a more realistic prospect for the foreseeable future. It suggests in turn that for the erasure of an unpleasant episodic (or semantic) memory to have the greatest chance of success, the procedure might need be performed as soon as possible after the event that gave rise to it, to avoid the "proliferation" of that particular recollection in the person's mind.

At present, the selective erasure of declarative memories in humans remains a speculative prospect. So far, researchers have succeeded in selectively removing a conditioned (implicit) fear memory in mice by ablating specific neurons overexpressing a protein called CREB[21]. Destroying neurons, however, is a rather extreme procedure, and alternative, less radical methods have since been discussed. Glannon, for instance, cites electrical stimulation of the brain (especially Deep Brain Stimulation)[22]. Most spectacularly yet, using the tools of optogenetics, researchers have demonstrated the reversible deactivation of a fear memory in rats[23]. Still, the caveats mentioned earlier about extrapolating results achieved in mice to humans, and about the difference between erasing implicit and declarative memories, apply again here. Further research will thus be needed to determine whether memories, including episodic ones of the kind that cause distress to many people, can also be selectively erased in humans. What can at least be said is that, *should* this indeed prove possible, it is quite straightforward to see how it could eliminate - at least in the short term - even non-pathological negative affect, unlike foreseeable forms of memory blunting. As the saying goes, "what you don't know won't hurt you".

In the following ethical discussion, I shall assume that it will ultimately become possible to selectively erase undesired human memories (including declarative ones) in a reasonably safe manner. I will further assume that it will also be feasible, when necessary, to couple memory erasure with the implantation of false memories. As indicated previously, such prospects raise a number of ethical questions, having to do with considerations like personal identity[24], autonomy[25], or a putative "duty to remember"[26]. In what follows, however, I shall focus on one specific question: namely, whether memory erasure would undermine the truthfulness of people's lives.

3. Memory erasure and the objection from "truthful living"

One justification commonly invoked by those who object to the prospect of erasing painful autobiographical memories using biomedical tools, especially when the memories in question do not give rise to genuine pathology (as in the case of *Eternal Sunshine*), involves an appeal to the value of authenticity, understood in the specific sense of "truthful living"[27]. This needs further spelling out, as it could involve two

distinct ideas. The first, main idea is that truthful living involves what we might call "clear-eyedness": living one's life on the basis of accurate beliefs about one's own personal circumstances, including the significant aspects of one's past. On this conception, the ethical problem with memory erasure is that one will then be "living a lie", no matter how pleasant that lie might be[28]. The second possible construal of truthful living stresses instead truthfulness *towards others*, that is, the accurate presentation of who we are to other people[29]. On this construal, memory erasure will threaten the truthfulness of our lives if it leads us to misrepresent who we are to others, even involuntarily, as a result of our self-induced inaccurate recollection of our own past. This concern about truthful living, in each of these two possible construals, will have even greater applicability in cases where memory erasure is coupled with the implantation of false memories - although the latter practice, taken on its own, need not be open to the charge, insofar as the person could still remember the implantation procedure and be able to identify the implanted memories as fictional.

The underlying assumption behind the idea that memory erasure threatens authenticity as clear-eyedness is that it is good to live one's life "truthfully" in that sense. Most will agree that it can be good to do so, at least in many cases, for instrumental reasons: for instance, because remembering our past with reasonable accuracy is important if one is to be able to successfully relate to others, as well as for personal growth (allowing us to learn from our past mistakes)[30]. Furthermore, truthfulness can arguably serve the cause of justice: a victim of bullying or sexual abuse who completely erased her memories of victimization would thereby lose one powerful incentive to actively oppose such practices. And were many victims to do the same, negative social effects might ensue, such as a weaker will on the part of society to fight victimization and hold perpetrators to account, and a diminished perception of the harmfulness of those practices (especially if it came to be expected that victims would erase their memories of what was done to them)[31].

Yet those sympathetic to the clear-eyedness objection will typically also hold that living "truthfully" can be good for its own sake, independently of its beneficial consequences - and even if it makes us less happy than we could otherwise be. They might for instance accept an "objective list" account of well-being that includes clear-eyedness among its components, possibly alongside pleasure, and one that allows the former value to outweigh the latter in at least some cases of conflict. Or they might rather view truthfulness as making for a life that is better in an impersonal, non-welfarist sense (e.g. more admirable). Robert Nozick's famous case of the "experience machine" is viewed by many, although not all, as lending support to the intuition that a less happy life can nonetheless be better than a happier one if it involves a greater connection to reality[32].

It is possible for the truthfulness objection to memory erasure to combine both senses of truthful living, insofar as part of the instrumental value of clear-eyedness might pre-

cisely be that it increases our ability to present ourselves accurately to others. True, as Adam Kadlac points out, people can sometimes achieve an accurate self-presentation in an unintentional manner[33]. For instance, a person's vanity might be fairly obvious to others even if that person herself were unaware of that trait. Clear-eyedness is therefore not a necessary condition of accurate self-presentation. Still, the prospect of memory erasure does suggest that the two are connected to some degree. Although we do not entirely depend on people's own reports about themselves and their personal history in order to form an accurate opinion of who they are, we nevertheless do need to rely on such sources to *some* degree. Independently fact-checking every self-report about someone's past is clearly not feasible. Given this, reductions in clear-eyedness resulting from memory erasure might in turn lead a person to present herself less accurately to others, despite her being just as sincere as before. As for the value of truthfulness in this second sense, it might again be taken to be either intrinsic or instrumental (or both). Accurately presenting who we are to others might be viewed as a virtue (a virtuous form of "authenticity"), as well as part of what it means to treat others with respect. But is also seems necessary to build trustful and meaningful relationships with others; clear-eyedness alone will not suffice if it is combined with deceit.

An important question here is how much force the truthfulness objection to memory erasure is supposed to have. It does not seem very plausible to construe it as a decisive objection to *all* forms of memory erasure, regardless of context. A more appealing view is that the value of truthful living provides a strong reason against erasing one's memories in any given case - provided that the memories in question are at least somewhat accurate. (Erasing a highly distorted recollection of some past event need not negatively impact the truthfulness of one's life.) Yet the reason in question remains a *pro tanto* one: that is, it can in principle be outweighed by competing considerations, such as the need to alleviate unnecessary suffering.

Many people sympathetic to the truthfulness objection will likely agree that the reason it yields against memory erasure is outweighed in this manner in cases where a traumatic memory gives rise to debilitating symptoms of PTSD, for example - even though they might then still regard the loss of truthfulness as a regrettable aspect of the best option available, yielding an incentive to try and develop truth-preserving alternatives of comparable effectiveness. By contrast, they might deny that the value of truthfulness can be outweighed when a memory, while painful and upsetting, does not result in a genuine mental pathology. The so-called therapy-enhancement distinction might be considered relevant here: relieving the suffering caused by traumatic memories at some cost to truthfulness might be ethically defensible, but this, one might add, does not extend to the "normal" negative affect associated with many bad memories.

To be clear, what the truthfulness objection so construed purports to show is that erasing bad memories is ethically problematic, at least when done for non-therapeutic purposes. This does not automatically imply that such interventions, even when of a

specifically non-therapeutic nature, should be *prohibited*. Insofar as it would restrict people's right to mental self-determination or cognitive liberty[34], a legal prohibition on memory erasure would need the support of very weighty considerations to be justified[35]. For instance, we should have good reasons to expect that allowing such procedures to be used would result in great social harm, perhaps because social cohesion and coordination would be seriously disrupted by the large number of people who chose to regularly rewrite their own inner stories as they saw fit. I shall suggest later on that this is not a very likely prospect. Alternatively, one might argue that at least some people have a "duty to remember", witnesses to a crime for example, and that they should barred from erasing the relevant memories, at least temporarily (say, until they had testified at an upcoming trial). While this issue is no doubt of considerable interest, I will not attempt to deal with it here. Rather, I shall focus on the relevance of truthfulness for the *ethics* rather than the regulation of memory erasure.

I will also be focusing on cases in which a person *freely* and *deliberately* chooses to have some of their memories erased. It is easy to imagine shocking scenarios in which such procedures are used in a coercive manner: for example, an authoritarian regime might try and force all the witnesses of a brutal act of repression to have their memories of the event erased. While living truthfully might plausibly acquire even greater importance in circumstances where one is faced with coercive pressures to forget, it seems fairly uncontroversial that such uses of memory erasure would be wrong, if only because they would represent unacceptable violations of cognitive liberty or mental integrity. I shall therefore leave coercive uses aside here. Similarly, accidentally erasing memories other than the target memory would clearly be among the potential near-term concerns that memory erasure would present, at least during the early human trials of the intervention. Such unintentional erasure would be undesirable in multiple respects, including truthfulness. Nevertheless, it also seems to leave much less room for ethical controversy than deliberate forms of erasure, which will be my focus here.

The truthfulness objection to memory erasure has been criticized from various angles. In what follows, I will look at two main lines of criticism.

4. Challenges to the truthfulness objection

4.1 Skepticism about the accuracy of autobiographical memories

The first line of criticism states that concerns about memory erasure and truthfulness are misguided insofar as they are based on an overestimation of the current reliability of our memories. In the words of Nada Gligorov, for instance, "research on human memory has revealed the precarious nature of remembrance. Memories are not a veridical representation of the past"[36]. In support of that statement, she cites studies by prominent memory researchers like Elizabeth Loftus, showing how memories of past

events can get distorted by a variety of factors, including the learning of new information (the phenomenon known as "retroactive interference"), people asking suggestive questions about the event, and the influence of stereotypes and expectations[37]. From this, Gligorov goes on to conclude that "we can then wonder to what degree our memory is ever a truthful record of the past even without the use of [memory modification technologies]"[38].

In a somewhat similar spirit, Marijn Kroes and Rain Livoja offer a rather damning assessment of the accuracy of our autobiographical memories, writing that "[w]hen witnessing distressing events such as the Challenger Space Shuttle explosion or 9/11, we are often very sure about the accuracy of our episodic memories, but in fact we accurately remember only around 30%"[39]. This estimate is based on some studies of so-called "flashbulb memories", unusually vivid memories of such momentous and emotionally charged events[40]. This could - although, I shall argue later, *should* not - be read as implying that any personal memory one might target for erasure has a 70% probability of being inaccurate[41]. Kroes and Livoja seem to echo Gligorov when they conclude that in the ethics debate on memory modification, "it is critical to realize that memories are not a veridical reflection of the past but serve to support adaptive responses and decision-making in the future"[42]. Similarly, Marcos Alonso Fernandez argues that human memory is fundamentally aimed at being "functional" (from the perspective of evolutionary goals like survival), rather than "truthful or extensive"[43].

The skeptical view, then, can be formulated as stating that, based on the available evidence, our autobiographical memories in their current state seem highly inaccurate. If so, the concern that erasing unpleasant autobiographical memories would threaten the truthfulness of our lives is mostly unwarranted, because for all we know, we are already living untruthfully. Admittedly, the truthfulness objection as formulated above did add the proviso that the target memories should at least be somewhat accurate in order to yield a reason against erasure, in which case the objection need not be strictly falsified by the said evidence of the fallibility of human memory. Indeed, it does not assert the presence of such a reason in cases of seriously inaccurate memories. Yet if few, if any, autobiographical memories (especially of the non-traumatic kind) happen to meet the accuracy condition, the truthfulness objection will at least become practically moot, as it will apply to very few real-life cases[44].

In fact, on such a skeptical view, it seems possible that in some cases at least, erasing a painful autobiographical memory might *promote* truthful living, insofar as the removal of a misleading memory will also remove the associated false beliefs about one's past. It is even possible - although perhaps not very likely - that a more pleasant memory that got implanted as a replacement for the original one might accidentally happen to be more accurate as well. But even when that is not the case, we might, according to the skeptic, simply end up trading one inaccurate memory for another equally inaccurate, yet more pleasant one. And surely, one might argue, if both are

equally untruthful, better to choose the more pleasant illusion. On the other hand, one might argue that if this skeptical view is correct, scientists and physicians would have an ethical obligation to warn people that the memories they were targeting for erasure (at least in non-therapeutic cases) were likely misleading. Indeed, such a warning might itself suffice to provide the psychological relief these people sought, by helping them realize that they most likely did not act shamefully, or were not victimized, in the way they thought they had - thus also sparing them the potential financial costs and unwanted side-effects of a memory erasing procedure. In other words, a skeptic might hold that there are indeed reasons to discourage people from resorting to memory erasure for non-therapeutic purposes; only that these reasons do not include the risk of undermining the truthfulness of our lives.

4.2 The value of truthfulness need not be decisive even in non-therapeutic cases

Even assuming that our autobiographical memories are not so suspect as to make concerns about truthfulness practically irrelevant, other critics will point out that the truthfulness objection (in its above formulation) concedes that truthfulness is not an overriding value: it can sometimes be outweighed by competing considerations, as in the case of people suffering from PTSD. On that basis, the objector might proceed to challenge the common assumption that such outweighing can only occur in cases in which the target memory gives rise to a full-fledged mental pathology. This assumption, they might contend, questionably presumes that the line separating "normal" unpleasant memories from pathological ones perfectly coincides with the line between cases in which truthfulness trumps all other considerations, and cases in which it does not. To this the objector might add that even if considerations of truthfulness are weightier than any competing ones in at least some cases of non-therapeutic memory erasure, this will only mean that the *best* thing to do in such cases is to retain the relevant memory. It need not entail that erasing the memory is ethically *impermissible*, as the truthfulness objection would have it. If, as commonsense morality seems to demand, we want to make room for *supererogatory* actions, actions that are the best to perform in the circumstances yet are not *required* in light of the expected costs to the agent, we should acknowledge their correlate, which philosophers have called "permissibly suboptimal" actions[45].

A first consideration that might be advanced against the presumption that truthfulness can only be outweighed in therapeutic cases of memory erasure involves the idea that such a procedure could potentially play a positive role in crime prevention. The prospect of a criminal wiping out their (non-pathological) memory of a serious crime she committed might intuitively strike many as very disturbing. One might worry, among other things, that this would make that person more likely to commit similar crimes in the future, since she could no longer be deterred by the recollection of her original

act, and any feelings of guilt it might occasion. However, some have argued, to the contrary, that erasing a criminal's memory of her act (presumably with her consent) could in fact help *reduce* the risk of recidivism. Matthew Liao and Anders Sandberg, for instance, write that "deliberate forgetting could decrease the likelihood of future crimes of this type, because remembering may make it easier to commit the crime in the future, since one has already done it before"[46].

Secondly, and more importantly, one might suggest that truthfulness can plausibly be outweighed by other values in much more mundane and common cases. For example, non-traumatic yet highly embarrassing memories of past mistakes can haunt a person for many years. Erasing these memories might help promote her psychological well-being. True, such memories can, at least sometimes, help a person learn and improve her behaviour in the future. Yet one might counter that this need not always be the case: besides their unpleasant phenomenology, such memories can also undermine a person's self-confidence and interfere with the quality of her decision-making[47]. By contrast, it seems that an unreasonably high self-esteem, based on an overly rosy picture of one's past, might sometimes help people accomplish more than they otherwise might have been able to. Even those who regard clear-eyedness as a component of human well-being are likely to also include values like happiness and achievement into their list. And, the objector will maintain, it is doubtful that clear-eyedness must necessarily, or even typically outweigh such values in cases where they conflict, regardless of the presence or absence of a mental pathology.

Furthermore, the objector might contend, even establishing that truthfulness outweighs all competing values in a given case is not yet sufficient to demonstrate the *wrongness* of erasing a memory. As some philosophers put it, an agent may sometimes have an "agent-favoring prerogative" implying that they are not *required* to do what would be best in a given situation[48]. Consider Walter Glannon's example of a young scientist who performs poorly while giving a presentation at a conference, making a serious factual error and finding himself unable to deal with criticism from experts in the audience who point it out. While this unpleasant experience does not, we may assume, cause the scientist to develop any full-fledged pathology, it nevertheless leaves him emotionally scarred, deeply embarrassed, and doubtful of his own abilities as a researcher[49]. Suppose that Glannon's scientist could, if he chose to retain the embarrassing memory of his past failure, use it to spur himself to work hard and make significant contributions to his field – thereby more than redeeming that failure. By contrast, we may assume that erasing the memory would lead him to achieve less, but to also experience much less anguish. His life in the latter case need not necessarily be happier overall than the more truthful one (perhaps the extra negative affect caused by the painful memory would be counterbalanced by the more intense joys accompanying his scientific triumphs), yet it would at least be emotionally more even-keeled.

In this variant of the scientist's example, memory erasure thus seems to yield an overall suboptimal outcome. Still, does it follow that it would be *wrong* for the scientist to erase that memory? Surely not, the objector might argue; thinking otherwise would seem unduly harsh. A more plausible verdict might be that the scientist would act *best*, and deserve praise, if he retained the memory and found in it the motivation to succeed, but that it would nevertheless be ethically permissible for him to erase it, and live a less impressive but emotionally more comfortable life.

In addition, to defend the permissibility of erasing memories even if we can foresee that this will mean presenting ourselves less accurately to others, one might reason along the lines of the Doctrine of Double Effect, and contend that this is relevantly different from deliberately deceiving others about our past. Even assuming the latter to be ethically problematic, this need not extend to the rewriting of unpleasant autobiographical memories in cases where we merely *foresee* that others might acquire false beliefs about us as a result, but do not *intend* this outcome (we might simply want to feel less burdened by the past).

To support this line of argument, suppose that Susan tends to feel nervous and anxious every time she has to engage in public speaking, yet that, by using anti-anxiety medication, she manages not to let these feelings affect her speeches. As a result, many members of her audience assume that she is a naturally fearless and confident speaker. While it might be ethically objectionable for Susan to go around falsely proclaiming that she is indeed such a person, it seems much less plausible to criticize her for being "untruthful" if she merely fails to take active steps to correct people's mistaken assumptions about her natural fearlessness – at least if we assume that no one ever asks her whether such assumptions are correct. This point arguably stands regardless of whether or not Susan can foresee that some audience members will make such mistaken assumptions. As Kadlac puts it, "most of us are not entirely forthcoming about everything in our lives, and that fact does not mean that we are all phonies"[50]. Similar remarks might be applied to the case of a person foreseeing that she is likely to unwittingly mislead others about her past as a result of erasing an unpleasant autobiographical memory.

5. Replies to the challenges

5.1 Countering skepticism about the idea of truthfulness

The skeptical challenge to the idea of truthfulness makes a valid point: the evidence does indicate that our memories, including autobiographical ones that we consider particularly significant, may be less accurate and more vulnerable to disruption than we usually recognize. Does this, however, imply that concerns about the potential impact of memory erasure on the truthfulness of our lives are moot, because in any given case, we will have good reason to suspect that the memory being targeted is inaccurate? I do not believe so.

To begin with, we may note that the skeptical argument seems to assume the memories targeted for erasure will be ones that a person has had for a long time, enough time for them to deteriorate and get contaminated by misleading information. While this may eventually become possible, I have mentioned the special challenge presented by the erasure of firmly established memories, which would potentially require removing a large number of interconnected ones, in addition to the original recollection. This would create an incentive to intervene early, shortly after the painful event occurred, so as to circumvent that challenge. Yet early intervention would weaken the grounds for skepticism about the accuracy of the target memory: our memory for recent events tends to be more accurate than our memory for the distant past. The skeptical challenge may therefore rest on questionable assumptions about the typical timing of memory erasure.

Let us nonetheless assume the feasibility of erasing even older memories. Even then, unsettling as the available evidence on the reliability of autobiographical memories may be, it does not support the strong skeptical challenge to the truthfulness concern. It is not entirely clear to me from which source Kroes and Livoja draw their estimate that "we accurately remember only about 30 %" of the important events in our personal history. They do cite an influential paper on "flashbulb memories" by Ulric Neisser and Nicole Harsch[51], which sought to assess the reliability of people's recollection of hearing about the disaster involving the space shuttle *Challenger* in January 1986. Neisser and Harsch were surprised by the degree to which people's recollections two and a half years after the event diverged from their testimony taken shortly after it. They identified a mean accuracy in recall among their respondents of 2.95 on a scale of 7. This corresponds to a mean accuracy level of about 42%, rather than 30%. However, such a figure would admittedly still seem to support a skeptical view of the reliability of autobiographical memories. And a recent survey of the views of memory researchers and other academics regarding the accuracy of memory for real-world experiences after 48 hours also found a median estimate of 40% – the estimate being even lower for memories two years after the event[52]. Furthermore, some studies of flashbulb memories do classify around 30% of subjects' recollections as "highly accurate" after 32 months. This includes Schmolck and colleagues' 2000 study of people's recollections of the O. J. Simpson trial verdict[53], as well as Neisser and Harsch's *Challenger* study, if we assume a "highly accurate" recollection corresponds to an accuracy score of 5 to 7.

Nevertheless, such data still do not warrant embracing the radical conclusion that more than two thirds of our autobiographical memories should be dismissed as inaccurate. First, one may question whether the available evidence does adequately support that conclusion. Some studies of flashbulb memories have thus found considerably higher mean accuracy scores: a study of memories of 9/11 by Kvavilashvili and colleagues thus arrived at a mean accuracy level of about 70% after 3 years[54]. Diamond and col-

leagues, who conducted the aforementioned survey of memory researchers, simultaneously present separate study results suggesting that "experts' intuitions about memory accuracy are overly pessimistic" - although they also acknowledge that memories of "more personally significant or social experiences" than those they studied are likely more vulnerable to distortion[55]. At present, the jury is still out on the overall accuracy level of personally significant autobiographical memories.

Secondly, even assuming a 30 or 40% figure accurately reflects that level, it is crucial not to misinterpret that figure. For one thing, it should not obscure the fact that there are significant individual differences in accuracy of recall among different participants in the relevant memory studies. A mean accuracy score is not a law that applies to each and every individual. Some of Neisser and Harsch's respondents, for instance, were more or less completely off the mark two and a half years after the *Challenger* disaster about how they first heard the news, while others, admittedly fewer in number, had largely (albeit not perfectly) accurate recollections. One takeaway from such studies, then, is that whether or not memory erasure will threaten truthfulness by diminishing the accuracy with which we remember our past might vary from person to person. While this might be less of a concern for individuals with generally unreliable autobiographical memories, it can have much greater relevance to others. Admittedly, virtually none of the respondents in those studies remembered their original experience with *perfect* accuracy. And at least some critics of the truthfulness objection to memory erasure do seem to define "veridical" memories as entailing perfect accuracy of recall. On such a stringent definition, hardly any autobiographical memories are veridical, and truthful living is indeed but a pipe dream. Yet this definition is surely excessively demanding. It does not match the characterization I have offered of truthful living as only requiring a *reasonably* accurate recollection of one's past.

This leads us to another important caveat: even assuming a particular person's autobiographical memories have an overall accuracy score of 30 to 40%, we cannot plausibly construe this as implying that any of these memories has a 60-70% chance of being completely false, and by contrast, a 30-40% chance of being fully accurate. The accuracy of memories is not such a binary issue, as skeptics themselves acknowledge when they remind us that memories are hardly ever perfect replicas of past experiences. A more plausible interpretation of such an accuracy score is that it indicates *how much* of a past experience a typical autobiographical memory accurately represents in that person's case. Furthermore, as Asher Koriat and colleagues have argued, a distinction needs to be drawn between *memory for gist* and *memory for detail*. There is good reason to think that we are significantly better at remembering the gist of a past experience than the details of it[56]. If so, we may reasonably assume that the 30-40% of their past our hypothetical person accurately remembers is likely to include at least much of the gist of what they lived through, whereas the erroneous part of their recollection is more likely to be about details.

Arguably, when it comes to truthful living, memory for gist is more important than memory for detail. That is not to say that the latter is irrelevant: we can certainly think of cases in which getting some details wrong can have momentous consequences (e.g. a victim of bullying misremembering who threw hateful epithets at her and thus feeling prolonged resentment towards the wrong person). However, it also seems clear that many affective responses depend on the gist rather than the specific details of an autobiographical memory. These include the pride elicited by the recollection of a significant personal achievement, and the suffering resulting from the memory of having been publicly humiliated. Glannon's scientist, for instance, will plausibly have what he needs to live truthfully, if he accurately remembers having made egregious factual errors in his presentation, been called out on it by experts in the audience, and found himself unable to properly respond to their criticism, even if he misremembers some of the details of what he said, or the identity of the particular people who criticized him. While accurately remembering those details would no doubt help maximally secure the benefits of truthful living, they do not seem strictly necessary for the scientist to have an accurate overall picture of what happened, and to experience emotions and adopt behaviours aligned with that picture (e.g. by ensuring he is much more careful with his empirical claims for his future presentations). By contrast, the scientist will likely go down the path of untruthfulness if he causes himself to remember instead that his presentation was error-free and widely acclaimed.

I say that he is "likely" to do so, because this will arguably depend on the details of the case. Indeed, we should note that truthfulness as clear-eyedness does not seem to solely involve remembering our past with reasonable accuracy. More generally, it entails seeing ourselves and our life circumstances in at least a reasonably accurate manner. Whether or not accurately remembering that failed presentation will lead the scientist, in the long run, to live his life in a more clear-eyed manner will depend on how exactly he responds to that recollection, and on the exact causes behind the failure. Suppose, first, that this professional hiccup is not the sign of a lack of ability on the part of the scientist, but rather evidence that he needs to work harder and be better prepared for his future presentations. Yet the scientist, feeling emotionally scarred by the experience, draws unduly pessimistic conclusions from it about his own capacities. As a result, he abandons his research activities and moves to a lower skill, clerical job instead, building his life on the false premise that he "just doesn't have what it takes" to be a good scientist. In this particular scenario, accurately remembering what happened turns out to have a negative impact on the scientist's overall level of clear-eyedness (and on his lifetime accomplishments).

Now suppose that, in a different scenario, the scientist draws the correct conclusions from his bad experience: he must become more conscientious if he wants to succeed at his chosen career. His accurate recollection will then be promoting clear-eyedness, whereas erasing it, and possibly substituting it with a pleasant memory of success,

would likely have the opposite effect (thereby preventing him from correcting course). The lesson from this is that accurately remembering past events is not a *sufficient* condition of truthful living (and its associated benefits). Nevertheless, it still seems that it is, at least in many cases, a *necessary* condition. While falsifying his memory of the failed presentation might protect the scientist against an unjustified loss of confidence in his own abilities, it would also deprive him of the evidence he needed to realize his work ethic was inadequate and had to be fixed.

In sum, the skeptical challenge to truthfulness does make valid points about the limits of the reliability of autobiographical memory, and about our tendency to be too quick to trust it. It also highlights the fact that concerns about memory erasure and truthfulness will have greater applicability to some people (those with at least a reasonably accurate autobiographical memory) than others. This in turn suggests a possible grain of truth in the radical skeptical conclusion that prospective candidates for memory erasure should be warned that the recollections they wished to erase were probably misleading. A more reasonable proposal might be that, before proceeding to erase a particular declarative memory (especially of a non-traumatic kind), such people should first be invited to test the general reliability of their autobiographical memory[57], as well as ensuring they had fact-checked the memory they wished to target. Those who turned out to score poorly on the test, and who could not corroborate their target memory with independent evidence, could then be encouraged to treat that memory with skepticism. This might be enough to provide the emotional relief they sought, rendering memory erasure unnecessary - although those among them who insisted on undergoing the procedure in spite of everything should arguably retain the freedom to do so.

Based on the current evidence, its merits notwithstanding, the skeptical view seems to go too far in its defiance of common sense assumptions about the reliability of memory. While memories do not require perfect accuracy to "support adaptive responses and decision-making in the future", they do have to be at least somewhat accurate. Being sharply disconnected from reality, including past reality, is rarely conducive to the fulfilment of one's goals, including the evolutionary goals of survival and reproduction.

5.2. Truthfulness does matter - but is not always required

As for the points that truthfulness-based reasons against memory erasure can in principle be outweighed even in cases of non-therapeutic use, and that such uses can be permissible even when truthfulness is not outweighed (so that erasure is then not the *best* available course of action), I believe they should both be conceded. The therapy-enhancement distinction does not seem to provide a fully reliable dividing line between ethically permissible and impermissible uses of such a procedure.

I will leave aside the proposal to use memory erasure as a method of crime prevention: besides the fact that there is currently little evidence suggesting the greater effective-

ness of such a method compared to alternative ones involving no interference with the offender's memories, that proposal would also need to be spelt out in more detail (for instance, would society have to ensure that the offender never be exposed to evidence of their crime after being released, and if so, how would this work?). Nevertheless, it is at least conceivable, for instance, that someone who occasionally erased memories of personal failures, such as a botched conference presentation, might in the long run achieve more thanks to her heightened self-confidence, and live a better life overall, than a person in similar circumstances who kept her memories intact, and consequently had a more accurate view of her own abilities, yet achieved significantly less as she felt much more timid and less inclined to take on big challenges. Furthermore, I agree that Glannon's scientist could not necessarily be said to have acted wrongly if he erased his memory of the failed presentation, resulting in a life lower in truthfulness and achievement, yet felt happier overall, or at least emotionally more even-keeled. Properly assessing his case would require fleshing it out further. For instance, it might become more plausible to describe him as acting wrongly if he resorted to memory erasure each and every time he encountered an unpleasant setback. Yet such a verdict would seem too harsh if his use of the procedure were a one-off event[58].

Nevertheless, this does not show the concern about memory erasure and truthfulness to be entirely devoid of merit. At least, if one is persuaded by the intuition that a truthful life is, all else being equal, a better or "higher" kind of life (admittedly not a universally shared view), one will have reason to think that the value of truthfulness would be decisive in many cases of non-therapeutic memory erasure. While the possibility that truthfulness might sometimes be outweighed by other considerations should be acknowledged, as discussed above, one should not be too quick, in any real-life case, to conclude without sound evidence that it is indeed outweighed, and that forfeiting truthfulness is the *only* way for the agent to secure the relevant competing goods. In many cases, there will likely be alternatives to erasure available: these will include working to "redeem" an unpleasant past by building a better future for oneself, or soliciting the help of friends, relatives, or a therapist to work through the pain caused by a particular memory, ultimately alleviating that pain (e.g. by altering one's perspective on what happened) or developing greater psychological resilience. The fact that attaining greater happiness or achievement through those alternative paths might be more challenging than simply erasing the memory does not mean that they are not available to someone. To establish that they are not, one must first make a serious effort to try them out. And in the (potentially numerous) cases where one's efforts are successful, it will indeed turn out that memory erasure was a suboptimal option (albeit not necessarily an impermissible one).

In addition, we may note that even in cases where the truthfulness-preserving alternatives just described were not viable options, or where the agent permissibly chose not to avail themselves of them, the value of truthfulness might still support *partial* over

complete memory erasure. More specifically, it might support erasing the *episodic* memory of the painful past experience, while ensuring that the person retained a *semantic* memory of what happened, rather than seeking to remove all recollection of it. Although this hypothesis would require further empirical corroboration, it seems plausible to assume that only remembering the general facts pertaining to such an experience would often be less painful for a person than the episodic memory she initially had. For instance, remembering that one had been viciously bullied by one's classmates in high school would likely be less upsetting than vividly remembering the bullying from one's first-person perspective as the victim. Similarly, simply remembering that one had badly botched a conference presentation, and failed to properly deal with criticism from the audience, should elicit less negative affect than a clear episodic memory of one's failure, which might get replayed over and over again in one's mind. While some truthfulness might still get lost in the shift from an episodic to a less fine-grained, purely semantic memory of a past experience, enough relevant information might still be retained to allow the person to remain reasonably in touch with her own past, and to learn from it.

Finally, while I agree that there is arguably an ethical difference between deliberately misleading other people about one's past, and merely foreseeing that one will do so as a result of memory erasure, the general claim that only the former practice, but not the latter, is ethically unacceptable is too sweeping to be plausible. Surely, whether or not it is permissible to mislead others, even unintentionally, as a result of rewriting one's inner story will depend on the details, including the magnitude of the rewriting. Suppose that John, after a series of meaningless sexual encounters, some of them unprotected, now feels ready to look for a serious relationship. He decides to erase the memories of these encounters, believing that doing so will make it easier for him to turn on a new leaf in his personal life and settle down. Even if John's motivation were not to deceive his future partner about his sexual history, his use of memory erasure might nevertheless be ethically problematic, especially if it put his partner at greater risk of catching a sexually transmitted disease. Therefore, different cases will warrant different assessments. Non-therapeutic uses of memory erasure that involved less serious forms of misrepresentation of one's past, and did not present a risk of harm to others, might be ethically permissible.

Even so, such uses of memory erasure might nevertheless still be prudentially undesirable on account of the false perceptions they would instil (both in the subject herself, and others). Erasing one's memory of the vicious bullying one had recently suffered might be permissible, yet it might later prevent the victim from forging a deeper connection with, say, their daughter, if she were to experience bullying herself. Removing the memory of an embarrassingly poor conference presentation, and replacing it with a false recollection of success, might lead a scientist to subsequently behave in ways that further damaged his reputation among those of his colleagues present at the event,

who might conclude that he did not simply fail to live up to the standards of excellence of his discipline on one particular occasion, but that he altogether lacks any understanding of what those standards are[59].

6. Practical constraints on deliberate forgetting

I have argued that memory erasure could, in certain circumstances, compromise the truthfulness of a person's life in normatively significant ways. I would, however, like to conclude this analysis on a cautiously optimistic note, by offering some reasons to believe that in addition to the technical challenges involved in successfully developing memory-erasing procedures for use in humans, the scope of such threats to truthfulness might be further limited by practical - and especially social - constraints on the possibility of deliberately forgetting, or actively rewriting, important life experiences. Indeed, attempts to make oneself *completely* forget that one had had such an experience would face multiple challenges, no matter how advanced the needed interventions might become.

First, memory erasure could either be limited to a particular target memory or memories, or also include the recollection of the erasing procedure itself. Knowing that one had erased a painful memory from one's mind would seem likely to make a person curious as to what that memory may have been. Could we reasonably expect people in such a situation to indefinitely resist the urge to find out the answer to that question? Although I can but conjecture here, this prospect does not strike me as very plausible. It seems more reasonable to anticipate that, in at least many cases, people would sooner or later choose to find out, whether from the provider of the memory-erasing procedure, or from acquaintances with the relevant knowledge. The idea that contractual or other legal provisions could be drawn prohibiting any third parties from revealing the relevant information at the person's request sounds somewhat fanciful. Such provisions would arguably infringe on important rights, such as freedom of expression, and perhaps more controversially, the right to access information extracted from one's own brain (even if it has since been deleted from it). Yet if someone were successful in assuaging their curiosity in such circumstances, they would at least end up restoring some semantic memories of the event they had initially sought to forget, thus partly defeating the purpose behind the original intervention.

It is not exactly clear, given the currently speculative nature of such an intervention, how erasing the memory of the erasing procedure could be made to work smoothly. Perhaps amnesia could be generated for the whole period over which the intervention took place, and the subject could be provided with an alternative, misleading explanation for the amnesic episode: say, that they had undergone a medical procedure requiring the use of a general anaesthetic. (The implantation of false memories to replace the original ones would be a more ambitious alternative.) However, besides requiring

the erasure of a number of different interconnected memories (e.g. the memory of deciding to undergo the erasing procedure, etc.), such a strategy would also involve deliberately and permanently deceiving the subject, rendering it ethically questionable, especially if practiced by medical professionals - even assuming the person had clearly consented to such deception beforehand.

Worse still, even if complete memory erasure (including the memory of the erasing procedure) were to ever become available, it is not clear that it would help avoid the aforedescribed problem of curiosity. In a society where that kind of intervention were widely available, people would have reason to suspect that they may have erased one or more of their memories, and caused themselves to forget it. This might again spur them to investigate, among the providers of such services and among their acquaintances, whether that was indeed the case. And again, it seems rather improbable that future societies would choose to enforce a requirement to deceive a person who had initially expressed the desire to forget about a particular event, but had subsequently changed their mind[60]. It is equally improbable that full cooperation, whether voluntary or coerced, could be secured from everyone familiar with that event to never take any steps that might remind the person of its occurrence - which would not only mean refraining from ever mentioning the event in their presence, but also concealing and perhaps even destroying any relevant evidence of it (photos, videos or other documents) they might possess.

Moreover, besides being unlikely to be implemented, such requirements would be ethically indefensible, given the unreasonable restrictions they would impose on other people's freedoms. We may occasionally have an ethical obligation not to remind someone of a painful experience they had somehow forgotten about: say, if we can expect that refreshing their memory would cause them renewed suffering without sufficient countervailing beneficial effects (including any intrinsic value truthfulness might have). Even so, enforcing such an obligation via *legal* requirements seems difficult to justify, except perhaps in exceptional circumstances, such as a nefarious agent deliberately seeking to trigger traumatic memories in a PTSD sufferer, for instance[61]. Outside of such extreme cases, recognizing a "right to forget" construed as the right not to be reminded of certain unpleasant experiences (rather than the more modest right to alter one's own brain using memory-erasing procedures) would again unacceptably curtail the basic freedoms of others, including freedom of expression.

Assuming the societies of the future do not choose to enforce such a problematic right to forget, whether via legislation or social expectations, we should therefore expect most uses of any prospective memory erasure technology to only involve the kind of *partial* erasure outlined above, which I have argued would present a lesser threat to truthfulness than complete erasure. Unless people wanted to wage a quixotic, lifelong battle against remembrance, by erasing their memory again every time they accidentally got reminded of the relevant past experience, they would typically have to agree to at least live with some semantic memories of that painful experience.

7. Conclusion

To sum up, the concern that selective memory erasure, should it get successfully developed for human use, risks diminishing the truthfulness of people's lives does not, as skeptics maintain, rest on clearly untenable empirical assumptions about the reliability of human memory. Nonetheless, valid questions remain about the proportion of cases in which such procedures would undermine truthfulness, given - among other things - the existence of substantial individual differences in the accuracy of autobiographical memories. I have also argued that it was reasonable to view truthfulness as both intrinsically and instrumentally significant, and thus as speaking against memory erasure in a number of cases. That said, it has to be acknowledged that considerations of truthfulness need not override all others, and that even when they do, this does not automatically imply the impermissibility of memory erasure.

Erasing episodic memories associated with trauma (or other mental pathologies), when solely targeting maladaptive implicit memories is not an option, is especially likely to be ethically permissible. Yet we can also think of cases where even non-therapeutic memory erasure seems permissible. If the truthfulness objection to memory erasure is construed as maintaining that there are no such cases, it is therefore unpersuasive. However, insofar as it entails that non-therapeutic erasure would often be suboptimal, even if permissible, given the foreseeable availability - in many cases - of truthfulness-preserving alternatives, the objection does make a valid point. Finally, I have argued that even though *full* memory erasure, if successfully carried out, would likely negatively impact truthfulness in many cases, this concern is mitigated - although not fully alleviated - by the practical constraints we can expect to bear on such interventions, making *partial* erasure a more plausible prospect. As long as future societies do not embrace a questionable right not to be reminded of certain past events, the availability of memory-erasing procedures thus need not seriously threaten the truthfulness of our lives.

Bibliography and Notes

Acknowledgments: I thank my research assistant Chan Ho-yee for helping manage the bibliography and notes. I also thank two anonymous reviewers for their helpful comments on a previous version of this paper. This research was supported by a GRF grant (14607918) from the Research Grants Council (RGC) of Hong Kong.

1. To simplify matters, the following discussion focuses on the editing of memories of unpleasant experiences. That said, memories of pleasant past events might also become candidates for editing, if they give rise to negative affect in the present.
2. Lavazza A, Memory-Modulation: Self-Improvement or Self-Depletion? Front Psychol. 2018;9. doi: ARTN 469 10.3389/fpsyg.2018.00469. PubMed PMID: WOS:000429254700001.

3. Liao SM, Sandberg A, The Normativity of Memory Modification. Neuroethics 2008;1:85-99.
4. Zawadzki P, Adamczyk AK, Personality and Authenticity in Light of the Memory-Modifying Potential of Optogenetics. AJOB Neurosci. 2021;12(1):3-21. Epub 2021/02/03. doi: 10.1080/21507740.2020.1866097. PMID: 33528319.
5. Curran H, Declarative and Nondeclarative Memory. In: Encyclopedia of Psychopharmacology [Internet]. Berlin, Heidelberg: Springer; 2014.
6. Stark SM, Stark CEL, Chapter 67 - Introduction to Memory. In: Hickok G, Small SL (eds), Neurobiology of Language. Amsterdam: Academic Press; 2016. pp. 845-6.
7. Ibid. p. 844.
8. Zawadzki P, Adamczyk AK, ref. 4.
9. E.g. Beckers T, Kindt M, Memory Reconsolidation Interference as an Emerging Treatment for Emotional Disorders: Strengths, Limitations, Challenges, and Opportunities. Annu Rev Clin Psychol. 2017;13:99-121. doi: 10.1146/annurev-clinpsy-032816-045209. PubMed PMID: 28375725; PubMed Central PMCID: PMC5424072.
10. Loftus EF, Creating False Memories. Sci Am. 1997;277(3):70-5. PMID: 9274041.
11. Saletan W, The Ministry of Truth: A Mass Experiment in Altering Political Memories. Slate [Internet]. 2010 28/08/2021. Available from: https://slate.com/technology/2010/05/slate-conducted-a-mass-experiment-in-altering-political-memories-were-you-fooled.html.
12. Liv N, Greenbaum D, Deep Fakes and Memory Malleability: False Memories in the Service of Fake News. AJOB Neurosci. 2020;11(2):96-104. doi: 10.1080/21507740.2020.1740351. PMID: 32228386.
13. Bernstein DM, Pernat NLM, Loftus EF, The False Memory Diet: False Memories Alter Food Preferences. In: Preedy VRea (ed.), Handbook of Behavior, Food and Nutrition. New York, NY: Springer Science+Business Media; 2011. pp. 1645-63.
14. Nash RA, Berkowitz SR, Roche S, Public Attitudes on the Ethics of Deceptively Planting False Memories to Motivate Healthy Behavior. Appl Cognit Psychol. 2016;30:885-97.
15. Hyman Jr IE, Billings FJ, Individual Differences and the Creation of False Childhood Memories. Memory 1998;6(1):1-20.
16. Ramirez S, Liu X, Lin PA, Suh J, Pignatelli M, Redondo RL, Ryan TJ, Tonegawa S, Creating a False Memory in the Hippocampus. Science. 2013;341(6144):387-91. doi: 10.1126/science.1239073. PubMed PMID: 23888038.
17. Phelps EA, Hofmann SG, Memory Editing from Science Fiction to Clinical Practice. Nature. 2019;572(7767):43-50. Epub 2019/08/02. doi: 10.1038/s41586-019-1433-7. PubMed PMID: 31367027; Gilbert F, Harris AR, Kidd M, Burnt in Your Memory or Burnt Memory? Ethical Issues with Optogenetics for Memory Modification. AJOB Neurosci. 2021;12(1):22-4. Epub 2021/02/03. doi: 10.1080/21507740.2020.1866104. PubMed PMID: 33528329.
18. Reconsolidation refers to the process by which consolidated memories that have been reactivated become susceptible to modification, once certain conditions are met.
19. Beckers T, Kindt M, ref. 9; Zawadzki P, Adamczyk AK, ref. 4.
20. Levy N, Changing One's Mind. The Ethics of Memory Erasure in Eternal Sunshine of the Spotless Mind. S&F. 2011;5:29-49.
21. Han JH, Kushner SA, Yiu AP, Hsiang HL, Buch T, Waisman A, Bontempi B, Neve RL, Frankland PW, Josselyn SA, Selective Erasure of a Fear Memory. Science 2009;323(5920):1492-6. doi: 10.1126/science.1164139. PubMed PMID: 19286560.

22. Glannon W, Brain Implants to Erase Memories. Front Neurosci. 2017;11:584. doi: 10.3389/fnins.2017.00584. PubMed PMID: 29114199; PubMed Central PMCID: PMC5660846.
23. Nabavi S, Fox R, Proulx CD, Lin JY, Tsien RY, Malinow R, Engineering a Memory with LTD and LTP. Nature 2014;511(7509):348-52. Epub 2014/06/05. doi: 10.1038/nature13294. PubMed PMID: 24896183; PubMed Central PMCID: PMCPMC4210354.
24. Glannon W, The Neuroethics of Memory : From Total Recall to Oblivion. Cambridge: Cambridge University Press; 2019.
25. Zawadzki P, Adamczyk AK, ref. 4.
26. Bublitz C, Dresler M, A Duty to Remember, a Right to Forget? Memory Manipulations and the Law. In: Clausen J, Levy N (eds), Handbook of Neuroethics. Dordrecht: Springer; 2015. pp. 1279-307.
27. Erler A, Optogenetic Memory Modification and the Many Facets of Authenticity. AJOB Neurosci. 2021;12(1):40-2.
28. The President's Council on Bioethics (U.S.), Beyond Therapy: Biotechnology and the Pursuit of Happiness. 1st ed. New York: ReganBooks; 2003; Grau C, "Eternal Sunshine of the Spotless Mind" and the Morality of Memory. J. Aesthet. Art Crit. 2006;64(1):119-33; Erler A, Does Memory Modification Threaten Our Authenticity? Neuroethics 2011;4(3):235-49. doi: 10.1007/s12152-010-9090-4. PubMed PMID: 22003375; PubMed Central PMCID: PMC3189328; Lavazza A, ref. 2.
29. DeGrazia D, Human Identity and Bioethics. Cambridge: Cambridge University Press; 2005; Kadlac A, The Challenge of Authenticity: Enhancement and Accurate Self-Presentation. J Appl Philos. 2018;35(4):790-808. doi: 10.1111/japp.12266.
30. A reviewer notes that some authors, such as Marya Schechtman, might also stress the importance of truthfulness for the constitution of one's own identity. Indeed, on Schechtman's approach, individual identity is defined by a person's self-narrative, provided that it is at least reasonably in touch with reality: Schechtman M, The Constitution of Selves, Ithaca. NY: Cornell University Press;1996. Given that I do not personally subscribe to such an approach to individual identity, I am not emphasizing, either, this particular argument for the significance of truthfulness.
31. Erler A, ref. 28. I thank an anonymous reviewer for highlighting the importance of this point.
32. Nozick R, Anarchy, State, and Utopia. Oxford: Basil Blackwell; 1974.
33. Kadlac A, ref. 29.
34. Bublitz JC, Merkel R, Crimes Against Minds: On Mental Manipulations, Harms and a Human Right to Mental Self-Determination. Crim Law and Philos. 2014;8:51-77.
35. Contra what I previously wrote, I no longer believe that medical professionals, for instance, can simply appeal to the importance of truthfulness (rather than other potential considerations, such as the therapy-enhancement distinction) as a justification for refusing to accede to a request for memory erasure: Erler A, ref. 28.
36. Gligorov N, Neuroethics and the Scientific Revision of Common Sense. Dordrecht: Springer; 2016.
37. Ibid. pp. 76ff.
38. Ibid. p. 83.
39. Kroes MCW, Livoja R, Eradicating War Memories: Neuroscientific Reality and Ethical Concerns. Rev Int Croix Rouge 2019;101(1):69-95, 78.
40. Neisser U, Harsch N, Phantom Flashbulbs: False Recollections of Hearing the News About Challenger. In: Winograd E, Neisser U (eds), Affect and Accuracy in Recall: Studies of 'Flashbulb' Memories. Cambridge: Cambridge University Press; 1992. pp. 9-31; Schmolck

H, Buffalo EA, Squire LR, Memory Distortions Develop Over Time: Recollections of the O.J. Simpson Trial Verdict After 15 and 32 Months. Psychol Sci. 2000;11:39-45.

41. To be clear, I am not claiming here that Kroes and Livoja intended their claim to be read in this manner, but only that it could be read as supporting the radical skeptical view I am presently discussing.
42. Kroes MCW, Livoja R, ref. 39.
43. Fernandez MA, Memory, Neuroscience and Memory Enhancement. Can J Bioeth. 2020;3(1):1-9. doi: https://doi.org/10.7202/1068759ar.
44. A reviewer points out that some memories can significantly affect a person's life even if they are not accurate. In such cases, one might argue that it is wiser not to rush into erasing the relevant (misleading) memory, and preferable to first embark on a course of therapy to understand its precise nature, how it came to be formed, and what its influence on the person's thinking and behaviour has been. Nevertheless, once this introspective work had been completed, it seems that the erroneous character of the memory would remove a key reason against erasing it.
45. E.g. McNamara P, Supererogation, Inside and Out: Toward an Adequate Scheme for Common-Sense Morality. In: Timmons M (ed.), Oxford Studies in Normative Ethics. 1. Oxford: Oxford University Press; 2011. pp. 202-35.
46. Liao SM, Sandberg A, ref. 3. p. 92.
47. Glannon W, ref. 24, p.132.
48. McNamara P, ref. 45.
49. Glannon W, ref. 24.
50. Kadlac A, ref. 29, p. 805.
51. Neisser U, Harsch N, ref. 40.
52. Diamond NB, Armson MJ, Levine B, The Truth Is Out There: Accuracy in Recall of Verifiable Real-World Events. Psychol Sci. 2020;31(12):1544-56. doi: 10.1177/0956797620954812. PubMed PMID: WOS:000597848300006.
53. Schmolck H, Buffalo EA, Squire LR, ref. 40.
54. Kvavilashvili L, Mirani J, Schlagman S, Foley K, Kornbrot DE, Consistency of Flashbulb Memories of September 11 Over Long Delays: Implications for Consolidation and Wrong Time Slice Hypotheses. J Mem Lang. 2009;61(4):556-72. doi: 10.1016/j.jml.2009.07.004. PubMed PMID: WOS:000272498600005.
55. Diamond NB, Armson MJ, Levine B, ref. 52.
56. Koriat A, Goldsmith M, Pansky A, Toward a Psychology of Memory Accuracy. Annu Rev Psychol. 2000;51:481-537,491-92.
57. Using tools such as the Autobiographical Memory Test (AMT).
58. Some might disagree, arguing that deliberate departures from clear-eyedness are always wrong. W. K. Clifford might be an example: see Clifford WK, The Ethics of Belief. In: Stephen L, Pollock F (eds), Lectures and Essays. London: Macmillan and Co.; 1886. Such a position, however, strikes me as too extreme in its absolutism.
59. Lavazza A, Moral Bioenhancement Through Memory-Editing: a Risk for Identity and Authenticity? Topoi 2019; 38 (1):15-27.
60. Another reason for society not to adopt a policy of withholding information about a person's use of memory erasure might be that, in a world in which people could fully erase their memories, we could expect unscrupulous parties to make spurious claims of erasure to further their own goals ("I definitely lent you money, but you made yourself forget about it!").
61. Bublitz C, Dresler M, ref. 26.