Published in final edited form as:

AJOB Neurosci. 2021 January ; 12(1): 52–54. doi:10.1080/21507740.2020.1866117.

Memory Deletion Threatens Authenticity by Destabilizing Values

Colton Hayse, Adina L. Roskies

Dartmouth College

Zawadzki and Adamczyk (2021) explore the consequences for authenticity of erasing and reinstating memories using optogenetics. Relying on empirically informed models of memory and nuanced models of identity, they consider the potentially dangerous threat to authenticity that modification of self-defining memories poses. They adopt a model of personality which is multifaceted, involving dispositional traits, characteristic adaptations such as motives and values, and narrative identity (McAdams 2013; McAdams and McLean 2013). Given the empirical evidence that semantic and episodic memory are functionally distinct, and that therefore each can persist despite damage to the other, the authors suggest that the survival of one's values in semantic memory might allow one to construct an identity coherent with these values, enabling a person to recreate an authentic self despite memory modification. Thus, they argue, despite the ability of optogenetic modification to remove and reinstate self-defining memories, such intervention does not pose a definitive threat to authenticity. We disagree.

Zawadzki and Adamczyk's argument rests on a diachronic model of authenticity borrowed from Pugh, Maslen, and Savulescu, in which authenticity involves coherence between values, identity, and rational beliefs:

It is through the lens of our evaluations, themselves developed in the light of our personal history and our stable, long-lasting characteristics and traits, that we are able to understand which of our features we want to be incorporated into our understanding of who we really are. (Pugh et al. 2017).

Most accounts of authenticity are *synchronic*, as is, for example, Frankfurt's (1971) account involving congruence between second-order and first-order desires. Synchronic accounts are vulnerable to an objection raised by Erler and Hope: people look to authenticity for guidance in endorsing desires or making decisions (Erler and Hope 2015). Therefore, a synchronic view of authenticity as endorsement has little to no practical value (Pugh, Maslen, and Savulescu 2017). A *diachronic* view of authenticity, on the other hand, can rest on a coherentist framework as laid out by Ekstrom, wherein authenticity requires coherence between one's "nexus of values" and one's "rational beliefs" (Ekstrom 1993; Pugh et al. 2017). Under a coherentist model of authenticity, values are enduring but not immutable, long-lasting though gradually updated by a "diachronic process of intelligible rational change" (Ibid. 2017). Pugh, Maslen, and Savulescu argue that we can "only intelligibly and justifiably change constituent parts of our true selves by appealing to other values that

we hold” (2017). For change to be authentic it must be rational, and any radical alteration of one’s values must be done in coherence with other values.

Given such a model, authenticity stems from harmony between values and narrative, and values are rendered stable in virtue of a rational, diachronic updating process. Zawadzki and Adamczyk argue that this stable harmony can be preserved following removal of even the most self-defining of memories, because values persisting within semantic memory can be gradually rediscovered within oneself after memory modification, as from a third-person perspective. This allows one to reconstruct an authentic identity following memory editing.

We caution that values are not as immutable as Zawadzki and Adamczyk suppose, even if they are functionally separate from episodic memories. Rather, given the particular coherentist model they promote, values must be *justified* to oneself. Justification is, in fact, the rational diachronic process to which the authors allude, by which values are rendered stable or unstable. If values are unjustifiable one revises them, until a framework is reached which *is* stable in the face of scrutiny. We argue that narrative identity plays a key role in this justification process, and that Zawadzki and Adamczyk underestimate this role, acknowledging the power of values to drive narrative yet too hastily dismissing the power of narrative to justify values.

As dispositional traits are endorsed or opposed based on one’s values, so values are justified or revised based partially on one’s narrative identity: one points to moments of personal experience to defend values, even those values which are semantically encoded. Our values may be regularly challenged, and if a challenged value is one whose grounding memory has been erased, it lacks any stabilizing defense. This is particularly evident when one must decide between two goods of equal moral status, where the sole consideration which normally justifies valuing one such good over another, rendering a decision between the two authentically one’s “own,” is personal identification with one good and not the other. To value one good over another objectively equal good would be arbitrary without a grounding in the personal experience of a self-defining memory. In the face of new challenges, or even the challenge posed by losing the grounding memory, one’s value system will have to arbitrate such a decision randomly, and, we argue, inauthentically.

Take, for example, the case of Elizabeth, borrowed by the authors from Erler (2011). Elizabeth works full-time for an anti-bullying organization as a result of her childhood experience as a victim of bullying. Elizabeth elects to work for this organization and not another social justice organization, say, one protecting victims of domestic abuse, despite there being no discernible difference in value from an objective standpoint between the two goods. Elizabeth works for an anti-bullying organization precisely because opposing bullying has specific value to her, undergirded by her identification with a particular state of victimization, something that relies upon recollection of vivid, traumatic personal memories of bullying. Zawadzki and Adamczyk argue that despite memory modification it is possible for Elizabeth to retain a “‘free-floating’ but persistent disposition” from which she can rediscover herself as she would discover someone other than herself, i.e. from a third-person perspective. They claim that through this third-person self-discovery she can eventually recreate an authentic identity as an anti-bullying activist.

Suppose Elizabeth's memories of being bullied were to be specifically targeted and modified by optogenetics. If Elizabeth were to learn of her disposition to work for an anti-bullying organization from a third-personal standpoint, she would analyze this disposition as she would anyone else's, with no special weight given to the plight of bullying victims because she has lost the ability to identify with the bullying-victim state. Yet it was this special consideration which justified her valuing anti-bullying activism over other types of activism. If asked why she works for an anti-bullying organization and not an antidomestic-abuse organization, she will fail to have an answer that can justify her choice. She may be able to say that she values it, due to persistent semantic memory, but she will be unable to identify with a lived experience that justifies *retaining and renewing* this value and thus prioritizing this work over other activities she considers good. Her value of anti-bullying activism will lose the rooting in first-personal experience which allowed her to justify that value to herself, and will become unstable in the face of other competing values. Yet values *must* be stable, i.e. justifiable, to be authentic, and so whatever value Elizabeth might replace her anti-bullying value with must be justifiable. But, we argue below, this is impossible. Therefore, to remove self-defining memories undermines one's evaluative structure by undermining the justification process itself.

A person who undergoes optogenetic memory modification of a self-defining memory now lacks that memory, and yet retains in semantic memory a value which implies that episodic memory; she therefore *knows that evidence which ought to inform her adjudication is missing from the justification process*. If a person excises a memory which formerly justified a core value, she may proceed by one of two equally inauthentic options. On the one hand, she may retain the now-unjustifiable value with which she no longer authentically identifies, because its foundation in a justificatory memory is now gone. On the other hand, she may adopt a new value which competes with the once-authentic value, but this adoption must overturn the former value without definitive experiential evidence that this ought to be done. In either case, what should be a "diachronic process of intelligible rational change" is supplanted by a vacuum of rationality. Consequently, her ability to justify *any value* in this sphere is undermined, and she is, by necessity, doomed to inauthenticity in this sphere (Pugh, Maslen, and Savulescu 2017). Zawadzki and Adamczyk acknowledge possibility of inauthenticity, but only in the case that memory modification damages semantic values *directly*. We counter that even if values are preserved in semantic memory, the removal of a self-defining autobiographical memory leaves semantic values unjustifiable and thus, by their own lights, renders choices that appeal to them inauthentic.

Zawadzki and Adamczyk argue that one might reconstruct an authentic narrative identity in the face of the modification of self-defining memories. We argue, in contrast, that modification of self-defining memories poses a greater threat to authenticity than Zawadzki and Adamczyk acknowledge. It may leave values unmoored in a manner that makes them impossible to justify, leaving them vulnerable (for example, to competition by other values), thus rendering once-authentic values unstable and the choices that depend upon them inauthentic. Therefore, even in the face of functionally separate semantic and episodic memory systems, optogenetic modification of self-defining memories poses a threat to authenticity.

FUNDING

This work was supported by NIH Brain Initiative Grant 1RF1MH117813.

REFERENCES

- Erler A 2011. Does memory modification threaten our authenticity? *Neuroethics* 4 (3):235–49. doi:10.1007/s12152-010-9090-4. [PubMed: 22003375]
- Erler A, and Hope T. 2015. Mental disorder and the concept of authenticity. *Philosophy, Psychiatry, & Psychology* 21 (3):219–32.
- Ekstrom LW 1993. A coherence theory of autonomy. *Philosophy and Phenomenological Research* 53 (3):599–616. doi:10.2307/2108082.
- Frankfurt HG 1971. Freedom of the will and the concept of a person. *The Journal of Philosophy* 68 (1):5–20. doi:10.2307/2024717.
- McAdams DP 2013. The psychological self as actor, agent, and author. *Perspectives on Psychological Science* 8 (3):272–95. doi:10.1177/1745691612464657. [PubMed: 26172971]
- McAdams DP, and McLean KC. 2013. Narrative identity. *Current Directions in Psychological Science* 22 (3):233–8. doi:10.1177/0963721413475622.
- Pugh J, Maslen H, and Savulescu J. 2017. Deep brain stimulation, authenticity and value. *Cambridge Quarterly of Healthcare Ethics* 26 (4):640–57. doi:10.1017/S0963180117000147. [PubMed: 28937346]
- Zawadzki P, and Adamczyk AK. (2021) Personality and authenticity in light of the memory-modifying potential of optogenetics. *AJOB Neuroscience*. 12 (1):3–21. doi:10.1080/21507740.2020.1866097 [PubMed: 33528319]