



## Identity, Immortality, Happiness: Pick Two

Shimon Edelman

Department of Psychology, Cornell University

[se37@cornell.edu](mailto:se37@cornell.edu)

*Journal of Evolution and Technology* - Vol. 28 Issue 1 – February 2018 - pgs 1-17

### Abstract

To the extent that the performance of embodied and situated cognitive agents is predicated on forethought, such agents must remember, and learn from, the past to predict the future. In complex, non-stationary environments, such learning is facilitated by an intrinsic motivation to seek novelty. A significant part of an agent's identity is thus constituted by its remembered distilled cumulative life experience, which the agent is driven to constantly expand. The combination of the drive to novelty with practical limits on memory capacity posits a problem. On the one hand, because novelty seekers are unhappy when bored, merely reliving past positive experiences soon loses its appeal: happiness can only be attained sporadically, via an open-ended pursuit of new experience. On the other hand, because the experiencer's memory is finite, longevity and continued novelty, taken together, imply eventual loss of at least some of the stored content, and with it a disruption of the constructed identity. In this essay, I examine the biobehavioral and cognitive-computational circumstances that give rise to this problem and explore its implications for the human condition.

Death is their fate, the gift of Ilúvatar, which as Time wears even the Powers shall envy.

– J. R. R. Tolkien, *The Silmarillion* (1977), ch. 1

### 1. A new angle on some old questions

Last October, on a trip to the Capitol Reef National Park, I went on what is regarded as one of the most rewarding canyon hikes in Utah: the Halls Creek narrows. Halls Creek follows the Waterpocket Fold south through the wilderness, eventually joining the Colorado at Lake Powell. About thirteen kilometers from the trail-head, it enters a spectacular wet narrows. For the next ten kilometers or so, the now perennial stream runs through a deep, meandering gorge, in places no wider than a couple of meters, hemmed in by soaring walls of red and white sandstone, which frame a narrow ribbon of ultramarine desert sky high above.

I spent the first day trudging with my backpack south along the stream bed and made my camp at the top of the narrows. In the morning, leaving the tent in place, I bypassed the gorge by an old trail running over a saddle east of the creek and entered the narrows from below. Every twist of the canyon held a surprise.

Will the sandstone cliffs past that pinnacle be white or wine-dark or banded? Will the next pool be too deep for wading, forcing me to swim? Will there be minnows in the pool? Will the quicksand claim my water sandals? Will there be monkeyflowers among the ferns on that wall? Will I come against an impassable obstacle – a chockstone or a fall – and be forced to turn back? I wished that the day would never end, but end it did. As the sun went down and I rounded the last bend of the gorge, my tent came into view and it occurred to me that the next day, instead of hiking back to the car, I could do another loop of the narrows.

Or could I? As Heraclitus pointed out long ago, you can't step into the same river twice. Perhaps more to the point, "the same you cannot step twice into the river" (Edelman 2012, 178). Let us assume that I kept collecting memories of experiences in the same manner that I now collect digital photographs, indexed and ready to hand, and stored in a repository whose capacity is unbounded (thanks to an unlimited supply of memory expansion kits). Could my self, when augmented in this manner – equipped with something like the Apple Photos app, but for entire episodic memories, and incrementally expandable – keep experiencing the thrill of novelty, time and again, indefinitely, in Utah or elsewhere?

It might be argued that, had I but world enough and time, and a brain wider than the sky,<sup>1</sup> that is, indefinite lifespan as well as bottomless memory, the accumulation of my experiences would eventually transform my old self, effectively making someone else happy in my stead (as suggested, notably, by Parfit, 1971). One way to avoid such a disappointing outcome could be to forgo indefinite expansion of memory, in which case I would sooner or later be forced to give up some of my past experiences, and with them those aspects of my self that they are part of. Or, of course, I could give up longevity, choosing instead to remain myself and to be happy, occasionally, for as long as the good life lasts.

The argument that I offer below suggests that the set of choices that define this predicament is quite general: it applies to humans as well as to any other type of cognitive agent with a similar evolutionary or design history. Specifically, I will make a case for the conclusion that one of three important possible attributes of an agent – indefinite lifespan, or integral identity predicated on cumulative memory, or the ability to be happy in the moment when experiencing novelty – must be given up.

My argument is built on an analysis of the interactions among these three attributes, each of which is key to understanding the human condition, and perhaps the condition of any future agents endowed with artificial general intelligence. Two of these attributes, happiness and memory, including the role of the latter in personal identity, have been extensively studied in a variety of disciplines, including psychology, biology, and evolutionary science, as well as philosophy. The third, indefinite lifespan or immortality, is an attribute that no humans possess, yet some consider desirable, and perhaps attainable through technology. All three are familiar themes of a multitude of literary works and many philosophical treatments, a few of which are referred to here.

In philosophy, there are ongoing debates concerning the relationship between memory and identity (Perry 2008b is a useful entry point into that literature), as well as between immortality and happiness (e.g., Thomson and Bodington 2014). A number of philosophers working in recent decades (e.g., Williams 1973; Momeyer 1988; Overall 2003; Thomson and Bodington 2014) have been moved to conclude that immortality should be shunned because it is arguably incompatible with happiness. As Christine Overall (2003, 165) put it,

At some point [...] the immortal person would have fully exploited all the capacities of his brain and body. He would arrive at a stage where his finite brain could not encompass any more and his finite body could not do or feel any more than they had already. In other words, the immortal's physical limitations would eventually place insuperable boundaries on his life prospects.

While I reach a similar conclusion, in my argument I attempt to extend the philosophical treatments of the

issues at the intersection of identity, immortality, and happiness by considering also a range of findings from psychology and other cognitive sciences, as well as evolutionary considerations. To sharpen the argument, I rely on the following construal of the three key concepts:

- (1) ***cumulative, integral identity***, as constituted by the long-term memory for personal experiences;
- (2) ***effective immortality***, extending to many times the normal human lifespan, but not concerned with cosmological time scales;
- (3) ***transient, situated happiness***, conceived of as the human emotional response to novelty and discovery in experiencing the world, and therefore more akin to joy than to life satisfaction (to use a standard distinction, stated in more detail below).

The rest of the paper is structured as follows. In section 2, I motivate my conceptualization of identity, immortality, and happiness and locate it relative to certain classical philosophical treatments of these matters. Section 3 then states the premises for my argument and the conclusions that they imply. Some possible objections to my framing of the concepts involved, as well as critiques of my conclusions, can be found in section 4. Finally, section 5 offers a personal view of the implications of my argument.

## **2. Background concerning the key concepts**

It should almost go without saying that the remarks contained in this section and the list of sources cited are not intended as (nor could they possibly be) a complete coverage of the history and the state of the art in the philosophy of memory and identity, or immortality, or happiness. Furthermore, when discussing these matters, and especially human memory and lifespan, I assume that technological progress will at some point make memory extension and extreme longevity feasible; quoting Paul Klee (2012, 188), “I have no desire to show [...] man as he is, but only as he might be.”<sup>2</sup>

### *2.1 Memory-based, cumulative, integral identity*

While it seems to be generally accepted that memory plays *some* role in forming a person’s identity,<sup>3</sup> philosophers disagree as to what that role might be (compare, for instance, Perry 2008a, parts II and III, to Perry 2008a, parts IV and V). My primary interest here is in subjective experience, because I would like to understand what factors contribute to my phenomenal selfhood (what it feels like to be me) and to be able to draw conclusions about happiness, which is of course subjective. Thus, I am less interested in arguments to the effect that memory is not a reliable logical criterion for identity (Shoemaker 1959, 873).<sup>4</sup> Nor am I concerned with putative logical attributes of identity such as “immunity to error from misidentification relative to the first person pronoun,” discussed in the context of memory by Shoemaker (1970, 270). In light of the contemporary scientific understanding of how the brain constructs the self (e.g., Gallagher 2000), and given how easily people can be induced to falsely recall autobiographical events (e.g., Loftus 2003), positing such immunity seems to be as disconnected from my present goals as striving to formulate a foolproof logical criterion for identity.

I am thus naturally driven toward the stance taken by Derek Parfit (1971, 8): “The alternative, for which I shall argue, is to give up the language of identity.” This choice leads to what he later formulated as the “Complex View” of personal identity (Parfit 1982) (as opposed to what he referred to as the more widespread, anti-reductionist “Simple View”). On Parfit’s psychological-reductionist view (which may be compared to the “psychological,” as opposed to “somatic,” approach discussed by Walker (2014, 165)), a person is gradually transformed by his or her experiences and by memories that result from those, so that the complex and dynamic cognitive structure that is the self changes over time, until eventually little or no

overlap may remain between the old self and a new one (see the diagram on page 24 of Parfit 1971).

Following Parfit, I assume that the accumulation of memories makes up a significant and enduring part of the self:<sup>5</sup> the narrative self (Dennett 1991; Neisser and Fivush 1994; Fivush and Haden 2003; Conway 2005), which is distinct from the momentary minimal or phenomenal self (Gallagher 2000). It is this life-long process of self-construction that John Keats referred to as “soul-making.”<sup>6</sup> I call this aspect of the narrative self the *cumulative integral identity* – my take on personal identity, which I introduced earlier and which serves as premise (6) for my argument, spelled out in section 3 below.

What if such cumulative memory were effectively infinite in its capacity? (Effective infinity, made possible by incremental “upgrades” as suggested in section 1, suffices for the present purposes.) When combined with extreme longevity or effective immortality (see section 2.2 below), the resulting unbounded build-up of retained experience records would, with time, change the integral identity of the host beyond recognition. This, of course, would make any further argument against immortality moot. And what if we require that the lifespan of the person be short enough so that there is not enough time for the integral identity to change too much as more and more memories accrue? This, too, would trivially force the argument’s conclusions, by ruling out immortality in the first place.

To avoid such a trivial resolution of the quandary concerning immortality, and to keep things interesting, our only recourse at this point with regard to memory is to assume that its capacity is finite. As long as no record of experience is ever expunged,<sup>7</sup> the identity of the host can be considered as integral in the sense defined here (modulo normal forgetting). Episode deletion will, however, necessarily result in a change to the host’s identity, of a complementary nature to the kind of change precipitated by episode accumulation. Whether or not a deletion ever becomes inevitable depends on an interplay of memory capacity, longevity, and happiness factors, as we shall see in section 3.

## 2.2 *Effective immortality*

Although Thomson and Bodington (2014) argue that absolute immortality, which requires an inability in principle for the agent’s existence to be terminated by whatever means, is the only kind that is really worth the name, my focus in this paper is on merely “effective” immortality, or extreme longevity, as defined earlier. Even, and perhaps especially, this more practical version provokes the question: is it desirable? This question is precisely the one that I will be in a position to address (in section 5), after examining immortality in the context of considerations of happiness and integral identity.

Effective immortality is commonly assumed, at least tentatively, to be desirable: “[. . .] there are considerable advantages (or at least purported advantages) to being uploaded, including immortality [. . .]” (Walker 2014, 175). It would seem that the categorically negative view of immortality on the part of Thomson and Bodington (2014) is due to their focus on an absolute, irreversible version of this concept. We already saw, however, that it is impossible to discuss the desirability of immortality separately from the nature of identity: thus, Overall (2003, 155) argues “[. . .] that two key concepts of being a person or self underlie debates about whether immortality is desirable and that which kind of person one chooses to be is related to whether one regards immortality as desirable.” With cumulative integral identity, as defined earlier, standing in for the concept of “person or self,” I can now proceed to discuss the concept of desirability.

## 2.3 *Transient, situated happiness*

It seems reasonable to assume that to be happy is desirable, but what does that mean? Of the many psychological aspects of happiness (e.g., Ryan and Deci 2001; Veenhoven 2003; Lyubomirsky et al. 2005), I am interested here exclusively in the hedonic variety, which (unlike the cumulative and sustained “life evaluation” component) is transient and situated, insofar as it is confined to the “here and now.” Note that

transient, situated happiness, or joy in the moment, happens to be an opposite of boredom, which, according to the now classical argument by Williams (1973), is the eventual fate, and perhaps the bane, of any human-like immortal. Arguably, a life that affords occasional experiences of joy is a good life and it can be pursued as a middle way between the unrealistic striving for constant joy (Edelman 2012) and a joyless existence which, if combined with immortality, does not sound at all enticing.

Now, a categorically negative valuation of immortality can be the consequence of an absolutist take on it, as per Thomson and Bodington (2014, 256):

[. . .] However improbable it might be that we would eventually exhaust even our most profound sources of meaning if we lived forever, if that is possible (and we see no credible way to deny it), then it will necessarily occur, given infinite time. This, we take it, is the deep point behind Williams' assertion that: "Nothing less will do for eternity than something that makes boredom unthinkable" (Williams 1993: 87).

Can a more nuanced position in the matter of immortality and boredom be formulated and defended? In this section, I have laid out an arguably more pragmatist take on all three concepts involved – cumulative integral identity, effective immortality, and transient, situated happiness. Can this framing of the key concepts, when combined with evolutionary, psychological, and computational (over and above philosophical) considerations, lead to a conclusion that is as unequivocal as that of Williams (1973) and others who hold immortality to be undesirable? In the next section, I lay out an argument to the effect that it can.

### **3. The argument, leading to a general principle**

The argument rests on several premises, drawn from evolutionary theory and cognitive science. Accordingly, it applies to any species of cognitive agents that are subject to realistic constraints (of the biological or engineering variety) and to evolutionary pressure and that are situated in a sufficiently complex environment. The statement of each of the six premises below is accompanied by a brief explanation and by a small selection of supporting references.

#### *3.1 The premises*

- (1) *Forethought confers evolutionary advantage.*

That forethought or foresight is evolutionarily advantageous has been argued in the past (e.g., Dennett 2003). The likely evolutionary and brain mechanisms underlying foresight and planning in primates, including humans, are discussed by Genovesio et al. (2014). Forethought, supported by various types of knowledge about how the world works, serves as *the* foundational theoretical concept in computational psychology (Edelman 2008, section 2.1.3; for an informal overview, see Edelman 2012, ch.2). The notion of predictive coding (which can be derived from the very general information-theoretic principle of free energy minimization; Friston 2010) has become a key explanatory tool in brain and cognitive sciences (e.g., Rao and Ballard 1999; Wacongne et al. 2011; Bar 2011; Clark 2013). The capacity for prediction may also contribute to the sense of agency and self-awareness (Gallagher 2000; Llinás and Roy 2009). Finally, it should be noted that the capacity for prediction is extremely general in that it embodies "a profound connection between the effective use of information and efficient thermodynamic operation: any system constructed to keep memory about its environment and to operate with maximal energetic efficiency has to be predictive" (Still et al. 2012, 120604-1).

(2) *Novelty-seeking boosts forethought in complex environments.*

In environments that are large enough to require exploration, or that change rapidly enough to require active tracking, an agent that relies on forethought may need to keep extending and revising the knowledge it has accrued in order to remain competitive. In such environments, a drive for novelty, or curiosity (Schmidhuber 2009), confers evolutionary advantage (e.g., Gottlieb et al. 2013).<sup>8</sup> Notably, intrinsically motivated learning (Baldassarre and Mirolli 2013) may outperform learning driven directly by “fitness” (that is, outcomes) over evolutionary time (Singh et al. 2010). Novelty-seeking through exploration in the service of forethought is thus an evolutionarily grounded answer to the question posed by Williams (1973, 93): “In general we can ask, what it is about the imaged activities of an eternal life which would stave off the principle [sic] hazard to which EM<sup>9</sup> succumbed, boredom.”

(3) *Happiness, like other emotions, evolves to regulate behavior, including exploration.*

The critical role of emotions in regulating cognition and behavior is well-documented (Roseman 2008; Winkielman et al. 2011; Lindquist et al. 2012; Inzlicht et al. 2015). With regard in particular to happiness, agent-based simulations of evolutionary dynamics (Gao and Edelman 2016a) and reinforcement learning (Gao and Edelman 2016b) suggest that a proper balancing of the two main components of happiness – hedonic (momentary; joy) and eudaimonic (strategic; life satisfaction) – is needed for optimal performance. In a foraging task, where exploitation must be combined with exploration, agents whose outcomes are linked to motivation via a well-tuned combination of momentary and longer-term “happiness” explore their territory and accrue fitness faster than agents that are motivated directly and exclusively by outcomes (Gao and Edelman 2016a, 2016b). When considered together with the evolutionary value of intrinsic motivation (Singh et al. 2010; see premise (2)) and with the cognitive and experiential benefits brought about by the feeling of awe stemming from exposure to striking natural vistas<sup>10</sup> (Rudd et al. 2012), these computational studies of happiness offer some intuition as to why traveling to new places (over and above experiencing abstract novelty) makes people feel good, for a while (hence the term *transient* situated happiness).

(4) *As a motive for exploration and a mediator of reward, honed by evolution, happiness is only effective when experienced episodically, rather than constantly.*

Because of hedonic adaptation (e.g., Lyubomirsky et al. 2005; Leventhal et al. 2007), no reward can remain of constant value if it is always available and no drive can remain equally effective if it is constantly present (Edelman 2012, ch.7).<sup>11</sup> It is important to realize that hedonic adaptation is merely a manifestation of the universal physiological phenomena of adaptation and habituation. Hedonic adaptation is a valuable trait: a species whose members are too easily satisfied with their performance and remain satisfied for too long after success is likely to be at a disadvantage relative to a species that is less prone to rest on its laurels. Indeed, evolutionary simulations suggest that agents in which hedonic states decay too slowly are more likely to go extinct (Gao and Edelman 2016a).

(5) *The memory capacity of any cognitive system is finite.*

This is one of the aspects of what philosophers of mind have called our “finitary predicament” (Cherniak 1986; Harman 1986).<sup>12</sup> In mammals, the capacity of long-term memory likely depends on the volume of the isocortex, which is thought to be “used up” as new memory traces are laid down over the lifetime of the individual (Merker 2004). Even if artificial expansion of memory capacity becomes possible, it should be noted that having a larger memory may negatively affect cognition. Because working memory, contrary to the popular notion, is not a dedicated “storage register” but a physically distributed emergent function that piggy-backs on long-term memory (Ma et al. 2014), a larger long-term capacity may interfere with flexible short-term use; this, in turn, may lead to a drop in fluid intelligence and to reduced cognitive performance (Wickelgren 1997; Gray et al. 2003; Conway et al. 2003). Moreover, certain purely computational

considerations suggest that larger memory capacity may result in poorer generalization (Tannenbaum, Yeshurun, and Edelman 2009), further impairing cognition.

(6) *Identity is constituted in part by cumulative life experience.*

Episodic memories accumulated over a lifetime – some repeatedly ruminated over, some distilled into concepts and narratives, many distorted to fit preconceptions, most pushed into the subconscious background – are such stuff as predictions are made on (Edelman 2008, ch.6). As discussed in section 2.1 above, they are also the stuff that makes up what I refer to as cumulative integral identity. With this final premise in place, I proceed to draw the conclusion.

### 3.2 *The conclusion drawn from premises (1)–(6)*

An agent whose performance depends on forethought (premise (1)), with a concomitant built-in drive to novelty (premise (2)), is faced with a problem. On the one hand, because novelty seekers who are bored lack situated happiness (premise (3)), and because happiness can only be attained sporadically, via an open-ended pursuit of new experience (premise (4)), reliving past positive experiences soon loses its appeal. On the other hand, because the experiencer’s memory is necessarily finite (premise (5)), longevity and continued accumulation of novel experiences, taken together, imply eventual loss of some stored content, and with it a disruption of the constructed identity (premise (6)). Consequently, for a situated agent, an effectively unlimited life span implies an eventual impossibility to both preserve the integrity of its identity and to be happy in the sense defined earlier. If such an agent opts for integral identity – and if it is forever safe, well-fed, personally attached, socially integrated, occupationally fulfilled, and satisfied with its achievements and status – its existence may include much contentment, but it will have little joy.

### 3.3 *The Rufus Trilemma*

The above conclusion points to a general, broadly applicable principle: THE RUFUS TRILEMMA.<sup>13</sup> Humans – or any other species of sentient agents, natural or artificial, with similar cognitive make-up and evolutionary history (or design parameters) – may attain at most any two, but not all three, of the following:

- integral identity;
- situated happiness;
- effective immortality.

In the next section, I consider some of the possible objections to the assumptions on which the Rufus Trilemma rests.

## 4. **Some possible objections**

The comments here are grouped into three sections, corresponding to the three clauses of the Rufus Trilemma, followed by a discussion of the limiting assumptions on which it rests. Note that this discussion complements and extends the comments that I made earlier, in section 2, on some of the relevant philosophy.

### 4.1 *Concerning integral identity*

**Integral identity in the long run need not be limited by capacity.** The popular notion that natural human memory capacity is “virtually unlimited” (as per, e.g., Michaelian (2011), cited above, who in turn cites R. A. Bjork) got a boost recently with the emergence of the first verified cases of hyperthymesia or highly superior autobiographical memory (HSAM; see the references in Patihis 2015). Like Funes the Memorious, in the eponymous story by Borges, people with HSAM remember in great detail an extremely

large number of their life experiences – an ability that may suggest that human memory capacity as such need not be an obstacle to integral identity even in the long run (especially if some quality of recollection could be traded off for even greater capacity). Such an interpretation of the HSAM findings is, however, too simple. For one thing, detailed investigation shows that HSAM is correlated with the one personality trait that is the most indicative of the malleability of personal identity: all 20 HSAM subjects available for the study (Patihis 2015) scored above age-matched controls on trait absorption, defined as “openness to absorbing and *self-altering* experiences” (Patihis 2015, 964, my italics).<sup>14</sup> If the cultivation of exceptional autobiographical memory brings with it a faster rate of personality change, “capacity” becomes not so much beside the point as simply meaningless, because memory can no longer be considered merely a storage device in the service of an immutable, integral agent.

**Integral identity in the long run is infeasible because of computational complexity, not capacity.** Even if human memory capacity as such is virtually infinite, it may still be limited by computational processing (e.g., retrieval) constraints. Thus, Michaelian (2011, 410) claims that “While the necessity of forgetting cannot be established by appealing to finite storage capacity, forgetting is indeed rendered necessary by the second aspect of the finitary predicament, limited computational resources.” The considerations on which this claim is based are, however, uninformed or mistaken. In particular, Michaelian fails to mention the well-known computational technique of locality-sensitive hashing (e.g., Andoni and Indyk 2008), which can make retrieval both extremely fast and capable of supporting similarity-based recall. Thus, if the integral identity clause should be dropped from the Rufus Trilemma, it is not because computational complexity renders it infeasible.

**Integral identity in the long run is trivially achievable through memory augmentation technology.** To a large and constantly growing extent, our memory storage and management needs are outsourced to external devices. Thus, my notebook computer holds tens of thousands of images documenting my personal life and travels, as well as thousands of academic papers that are relevant to my work; the sum total of human knowledge that is distributed throughout the internet at large is also easily accessible to me. Such artificial augmentation of memory (Burkell 2016) would seem to obviate arguments rooted in human neuroscience. However, as I noted in section 2.1 (following Parfit 1971), unchecked accumulation of personal memories would eventually obliterate the person’s original identity. To that, I might now add that, much as external memory is useful to us, as a repository of personal experiences it may have the wrong phenomenology. Specifically, computational considerations suggest that a digital system such as an external memory, even if closely coupled to a brain (e.g., via an implant), could never be integrated into its dynamics so as to be experienced naturally (Fekete and Edelman 2012; Fekete et al. 2016). Until an analogue (as opposed to digital) substrate for external memory becomes available, whose dynamics, moreover, would feel natural to us, the finite memory predicament will remain an impediment to unbounded integral identity in the long run.

**Integral identity in the long run is undesirable because forgetting is essential to being human.** Much research in cognitive sciences suggests that memory is constructive (e.g., Koriat and Goldsmith 1996; Glenberg 1997) and that forgetting is essential to its fulfilling its function, which is supporting decision-making (e.g., Richards and Frankland 2017). On a more philosophical note, Connerton argues persuasively that forgetting is “constitutive in the formation of a new identity” and that “what is allowed to be forgotten provides living space for present projects” (2008, 63). (Burkell (2016) invokes this argument in support of the notion that people should have the power to manage their online “footprints.”) A related notion, expressed by Mach (1886, 13), is that continuity, not integrity, is central to personal identity:

The ego is not a definite, unalterable, sharply bounded unity. None of these attributes are important; for all vary even within the sphere of individual life; in fact their alteration is even sought after by the individual. Continuity alone is important.

Indeed, Mach (1886) deplores clinging to the ego and envisages a future in which

We shall [. . .] no longer place so high a value upon the ego, which even during the individual life greatly changes, and which, in sleep or during absorption in some idea, just in our very happiest moments, may be partially or wholly absent.

As I already pointed out, this view of the nature of the self, just like that of Parfit (1971), can be seen as an argument, not against integral identity, but against extreme longevity (let alone immortality). Furthermore, just as one may “rebel” against death (e.g., de Unamuno 1972; Momeyer 1988), some of us deplore forgetting (Edelman 2014). In particular, I do not think I can credibly claim to have been happy if my recollection of any experience is soon lost, as in anterograde amnesia (e.g., Mauguière and Corkin 2015), or forcibly discarded to make room for a new one.<sup>15</sup>

#### *4.2 Concerning situated happiness*

Of the two main components of happiness, eudaimonic and hedonic, the second clause of the Rufus Trilemma has to do with the latter. The following discussion points pertain specifically to what I previously called transient situated happiness.

**Happiness requires identity, as a matter of logic.** Inasmuch as what matters here and now, for me, is *my* happiness rather than some kind of abstract or disembodied variety (as brought about, for instance, by a trance-like experience; e.g., Fischer 1973; Lebedev et al. 2015), it would seem that happiness is logically predicated on identity.<sup>16</sup> If that is the case, then someone who picks happiness would be logically required to also pick identity (but not the other way around), thereby upsetting the three-way symmetry of the trilemma. However, for this concern to be neutralized, it suffices that merely a modicum of personal identity exists for happiness to be anchored to; in contrast, in the Rufus Trilemma, the identity in question is integral – not merely a sliding window that retains the recent experiences while letting go of the old ones, but rather the cumulative sum total of one’s experiences over the lifetime.

**Happiness is not something that humans can reasonably expect to hold on to.** The biobehavioral, and therefore necessarily evolutionary, take on happiness (Buss 2000; Nesse 2004) and psychophysiological phenomena such as hedonic adaptation, discussed in section 3 are sometimes used to deflate the exaggerated expectations underlying the “pursuit of happiness” that characterizes the Western (and especially American) outlook. The deflationary approach can, however, be reconciled with the apparently universally human desire to be happy, by focusing on pursuit and forgoing the clinging that tends to accompany attainment (see Edelman 2012, and the many references therein). Such pursuit must be at least occasionally rewarded if it is to make the pursuer happy, leading us right back to the second clause of the Trilemma.

**Happiness can be achieved by means other than open-ended experiential novelty.** This sentiment has often been voiced by ancient philosophers who elevated equanimity over aspirations (as in “The satisfaction of contentment is an everlasting competence”; Lao Tze (Laozi) 1904). However, the vaunted equanimity is all about eudaimonia, not transient situated happiness. For the evolutionary reasons already discussed, we are doomed to seek, in new experiences, fleeting happiness that cannot be found in new possessions<sup>17</sup> or new acquaintances,<sup>18</sup> let alone in ruminating over the past.<sup>19</sup>

**Happiness should be taken under full control and subjected to reason.** It may be possible to train oneself to forgo both clinging and striving (and therefore both the pursuit of happiness and the happiness of pursuit), which are perceived as fraught with disastrous consequences by Buddhists (e.g., Scharfstein 1998; Siderits 2007). Indeed, technological means for “reaching in” and modifying the relevant physiological drives and reward settings may become available in the future.<sup>20</sup> As Williams (1973, 95) noted, “One might make the immortal man content at every moment, by just stripping off from him consciousness which would

have brought discontent by *reminding* [my emphasis] him of other times, other interests, other possibilities.” I have pointed out elsewhere (Edelman 2012, 67) that doing so would mean becoming something other than human, setting oneself well apart from the rest of humanity.

#### 4.3 Concerning immortality

**Immortality is not technically possible, so why fret over purely hypothetical questions?** (This question is worth dwelling on for a moment, even though my argument involves extreme longevity rather than absolute immortality.) Natural longevity in animals is indeed limited, probably by more than one factor. In humans, in particular, “the maximum lifespan [. . .] is fixed and subject to natural constraints” (Dong et al. 2016, 257). Intriguingly, in primates the best predictor of longevity is the size of the isocortex, which, as mentioned earlier, may correspond to memory capacity (Merker 2004). What about transcending the biological constraints entirely, by moving the contents of one’s mind to an artificial system? On the one hand, because the vehicle/content (or hardware/software) distinction does not apply cleanly to biological cognitive systems, copying a mind would necessitate copying its anatomy and physiology down to the molecular level.<sup>21</sup> On the other hand, a purely functional simulation of a mind, implemented in a digital substrate (as in “uploading”; Walker 2014), would necessarily possess radically different low-level dynamics, resulting in a major disruption (and likely a total obliteration) of the phenomenal self (Fekete and Edelman 2012). However, while the latter (digital simulation) route is unsuitable in principle, the former one – recreating the brain/mind from the bottom up in an artificial analogue substrate – is merely an engineering challenge, which cannot be dismissed out of hand. Thus, the immortality clause of the Rufus Trilemma still stands.

**Immortality should/will hold no appeal for progressive humans.** This last, normative point has been made by Mach in the context of his discussion of the mutability of the ego, from which I quoted earlier. The critical step, according to Mach, is to recognize that the self is in any case impermanent; “we shall then be willing to renounce individual immortality [. . .]” (Mach 1886, 13). Public opinion on these matters varies. Following a survey of the popular culture, Vidal (2016, 667) remarks that “Ambivalence is perhaps inherent to issues of longevity and, a fortiori, immortality. We may be youth-obsessed, and afraid of aging and death, but that does not necessarily make us wish to live forever.” Still, at least some of us would be unwilling to give up on prospects for immortality just because it seems to others selfish, unnatural, immoral (by religious standards; see Buben 2017) or simply tedious.

#### 4.4 Concerning the scope of the assumptions and the import of the argument

As stated up front in section 1, the argument developed in this paper deals not with identity, immortality, and happiness in general, but rather specifically with cumulative integral identity, effective immortality, and transient situated happiness, as defined here. Given the relatively restricted scope of these concepts, the conclusions they afford are also less than general. As some of the foregoing discussion suggests, relaxing the restrictions – for instance, by settling for (or, in a design context, opting for) fluid and mutable rather than cumulative and integral identity – makes the argument moot. It is conceivable that certain other changes in the premises could modify rather than obviate the argument. For instance, there are good philosophical arguments for the relevant concepts being “matters of degree” (Parfit 1982, 228) and there is good psychological and biological evidence, as well as computational reasons, for aspects and faculties of the mind/brain being rarely if ever categorical (Edelman 2008). This suggests that the conclusions of my argument may also be a matter of degree and will hold in proportion to the degree of integrality, happiness, and longevity, assumed or allowed. While a restating of the Trilemma in appropriately graded terms awaits future work, in its present form it applies at least to some of us or some of our progeny, natural or artificial.

## 5. Concluding remarks

Insofar as one must eschew one of the three great gifts – identity, happiness, and immortality – so as to gain the other two, the three alternatives that we humans will face in some not-too-remote and technologically quite plausible future are best labeled by what each of them requires us to give up: integral memory, in favor of happiness and immortality; or situated happiness, in favor of memory and immortality; or effective immortality, in favor of happiness and memory. Given how deeply personal this choice would have to be,<sup>22</sup> it would be presumptuous to offer any kind of general advice here. It may be noted, however, that the great gifts, and the implications of choices that involve them, have long been explored in the literary and other arts. A notion related to these gifts – that to make the most of the grand gift of life, one must remember, and be able at will to relive, *everything* – has also been given a literary treatment (Edelman 2014). It seems to me that the choice implicit in this notion amounts to admitting that the greatest gift to humans may be not life, about which those who have been born have no say anyway, but death – “the gift of Ilúvatar, which as Time wears even the Powers shall envy.”

## Acknowledgments

I thank Eric Dietrich for timely encouragement and for pointing out to me the connection between happiness and identity; Oren Kolodny for a discussion of memory, happiness, and hiking; two anonymous reviewers for useful comments; Russell Blackford for clear, detailed, and extremely helpful editorial suggestions; and Itamar Edelman for helping me focus on things that matter.

## Notes

1. With apologies to Andrew Marvell and Emily Dickinson.
2. Here is the quote from Klee (2012, 188) in its entirety: “If I had wished to represent the man ‘as he is,’ I should have required so bewildering a tangle of lines that a pure treatment of the element would have been out of the question; there would only have been an unrecognizable blur. Besides, I have no desire to show this man as he is, but only as he might be.”
3. An extreme view is that memory is “[. . .] the history that writes the individual, the narrative that creates the continuity called the person. If memory constitutes the person, then to remember what was is to be aware of who is, and to remember everything would be to see the person in its manifest fullness” (Lopez 1992, 35).
4. Shoemaker (1959, 873) writes: “Whether or not memory is a criterion of personal identity, it is not *the* criterion. [. . .] And while it is true that one does not use bodily identity as a criterion of personal identity when one says on the basis of memory that one did something in the past, this is not because one uses something else as a criterion, but is rather because one uses no criterion at all.”
5. And, on some philosophical accounts (e.g., Goodman 1978; Putnam 1982), also of the very world that the person inhabits.
6. Journal-letter to George and Georgiana Keats, April 21, 1819 (excerpted in Strachan 2003).
7. What I have in mind here is the kind of forceful deletion of episodic records that are central to the person’s emotional life, as depicted in the film *Eternal Sunshine of the Spotless Mind* (2004; written by Charlie Kaufman and directed by Michel Gondry); this is to be distinguished from normal human forgetting. More about this in section 4.

8. In the context of the human expansion out of Africa, it is interesting to consider how in various populations the distance from the cradle of civilization correlates with the frequency of an allele associated with openness to risk-taking, found in the gene that codes for the DRD4 dopamine receptor (Chen et al., 1999; Matthews and Butler 2011).
9. EM, or Elina Makropulos, is the immortal protagonist of a play by Karel Čapek, discussed by Williams (1973).
10. Such as those revealed to John Wesley Powell (2003, 397), “sublimity [. . .] never again to be equaled on the hither side of Paradise.”
11. Here’s Walt Whitman on transient situated happiness: “You but arrive at the city to which you were destin’d – you hardly settle yourself to satisfaction, before you are call’d by an irresistible call to depart” (*Leaves of Grass: Song of the Open Road* 82:11).
12. Michaelian (2011, 407) states it as follows: “Because her storage capacity is finite, if a human being were to live for a sufficiently long time, she would eventually run out of capacity.” The discussion that follows is, unfortunately, disconnected from recent and contemporary research in the psychology and physiology of human memory, resulting in mistaken claims such as that human memory has “virtually unlimited capacity.”
13. Marcus Flaminius Rufus is the central character in the 1947 short story *The Immortal* by Jorge Luis Borges (reprinted in Borges 1970). Rufus seeks and gains immortality, only to become disillusioned and weary with his interminable life. Mary Wollstonecraft Shelley’s story *The Mortal Immortal* (1829) is likewise built on the premise of immortality becoming unbearable. For additional examples from literature and film, see Vidal 2016.
14. Interestingly, the results of (Patihis 2015) also highlight the usefulness of memory for prediction. On item #28 on the Tellegen Absorption Scale (“I often know what someone is going to say before he or she says it”), HSAM subjects scored at .416, compared to controls, who scored at .272 ( $p < 0.01$ ). Likewise, on item #20 (“Things that might seem meaningless to others often make sense to me”), HSAM subjects scored at .594, compared to controls’ .182 ( $p < 0.001$ ).
15. Such purging of the mind’s records of experience would amount to acting bulimic, like Emperor Claudius (Crichton 1996), but with regard to memory.
16. This argument has been suggested to me by Eric Dietrich.
17. Experiential acquisitions are known to bring about more enduring satisfaction than material ones (Van Boven and Gilovich 2003); they are also more constitutive of the self (Carter and Gilovich 2012).
18. For a study of the role of friendship in happiness, see Saldarriaga et al. (2015). Momeyer (1988, ch.III) discusses the repercussions for social attachment and happiness of the hypothetical scenarios in which a person alone, a group of people, or everyone attains immortality.
19. Killingsworth and Gilbert (2010), whose subjects reported being less happy when caught with their mind wandering as opposed to when being focused on the task at hand, quote Keats: “Where but to think is to be full of sorrow” (“Ode to a Nightingale,” line 27).
20. As they have in science fiction, e.g., in Greg Egan’s novel *Permutation City* (1994).

21. I gloss over the fact that the question of the proper level here remains essentially unresolved (Fekete et al. 2016).

22. It would be interesting to see how people's choice in this matter correlates with their personality traits.

## References

Andoni, A., and P. Indyk. 2008. Near-optimal hashing algorithms for approximate nearest neighbor in high dimensions. *Communications of the ACM* 51: 117–122.

Baldassarre, G., and M. Mirolli, eds. 2013. *Intrinsically motivated learning in natural and artificial systems*. Berlin: Springer.

Bar, M., ed. 2011. *Prediction in the brain*. New York: Oxford University Press.

Borges, J. L. 1970. *The Aleph and other stories, 1933–1969*. Trans. Norman Thomas di Giovanni in collaboration with the author. New York: E. P. Dutton.

Buben, A. 2017. Personal immortality in transhumanism and ancient Indian philosophy. *Philosophy East and West*. In press.

Burkell, J. A. 2016. Remembering me: Big data, individual identity, and the psychological necessity of forgetting. *Ethics and Information Technology* 18: 17–23.

Buss, D. M. 2000. The evolution of happiness. *American Psychologist* 55: 15–23.

Carter, T. J., and T. Gilovich. 2012. I am what I do, not what I have: The differential centrality of experiential and material purchases to the self. *Journal of Personality and Social Psychology* 102: 1304–1317.

Chen, C., M. Burton, E. Greenberger, and J. Dmitrieva 1999. Population migration and the variation of dopamine D4 receptor (DRD4) allele frequencies around the globe. *Evolution and Human Behavior* 20: 309–324.

Cherniak, C. 1986. *Minimal rationality*. Cambridge, MA: MIT Press.

Clark, A. 2013. Whatever next? Predictive brains, situated agents, and the future of cognitive science. *Behavioral and Brain Sciences* 36: 181–204.

Connerton, P. 2008. Seven types of forgetting. *Memory Studies* 1: 59–71.

Conway, A. R., M. J. Kane, and R. W. Engle. 2003. Working memory capacity and its relation to general intelligence. *Trends in Cognitive Sciences* 7: 547–52.

Conway, M. A. 2005. Memory and self. *Memory and Language* 53: 594–628.

Crichton, P. 1996. Were the Roman emperors Claudius and Vitellius bulimic? *International Journal of Eating Disorders* 19: 203–207.

de Unamuno, M. 1972. *The tragic sense of life in men and nations*. Princeton, NJ: Princeton University Press.

- Dennett, D. C. 1991. *Consciousness explained*. Boston, MA: Little, Brown & Company.
- Dennett, D. C. 2003. *Freedom evolves*. New York: Viking.
- Dong, X., B. Milholland, and J. Vijg. 2016. Evidence for a limit to human lifespan. *Nature* 538: 257–65.
- Edelman, S. 2008. *Computing the mind: How the mind really works*. New York: Oxford University Press.
- Edelman, S. 2012. *The happiness of pursuit*. New York: Basic Books.
- Edelman, S. 2014. *Beginnings*. Kindle ed. Pennsauken, NJ: BookBaby.
- Egan, G. 1994. *Permutation city*. London: Orion.
- Fekete, T., and S. Edelman. 2012. The (lack of) mental life of some machines. In *Being in time: Dynamical models of phenomenal experience*, ed. S. Edelman, T. Fekete, and N. Zach, 95–120. Amsterdam: John Benjamins.
- Fekete, T., C. van Leeuwen, and S. Edelman. 2016. System, subsystem, hive: Boundary problems in computational theories of consciousness. *Frontiers in Psychology* 7: 1041.
- Fischer, R. 1973. A cartography of the ecstatic and meditative states. *Leonardo* 6: 59–66.
- Fivush, R., and C. A. Haden. 2003. *Autobiographical memory and the construction of a narrative self: Developmental and cultural perspectives*. Mahwah, NJ: Erlbaum.
- Friston, K. J. 2010. The free-energy principle: A unified brain theory? *Nature Neuroscience* 11: 127–38.
- Gallagher, S. 2000. Philosophical conceptions of the self: Implications for cognitive science. *Trends in Cognitive Sciences* 4: 14–21.
- Gao, Y., and S. Edelman 2016a. Between pleasure and contentment: Evolutionary dynamics of some possible parameters of happiness. *PLoS One* 11(5), e0153193.
- Gao, Y., and S. Edelman. 2016b. Happiness as an intrinsic motivator in reinforcement learning. *Adaptive Behavior* 24: 292–305.
- Genovesio, A., S. P. Wise, and R. E. Passingham. 2014. Prefrontal–parietal function: From foraging to foresight. *Trends in Cognitive Sciences* 18: 72–81.
- Glenberg, A. M. 1997. What memory is for. *Behavioral and Brain Sciences* 20: 1–55.
- Goodman, N. 1978. *Ways of worldmaking*. Brighton, Sussex: Harvester.
- Gottlieb, J., P.-Y. Oudeyer, M. Lopes, and A. Baranes. 2013. Information-seeking, curiosity, and attention: Computational and neural mechanisms. *Trends in Cognitive Sciences* 17: 585–93.
- Gray, J. R., C. F. Chabris, and T. S. Braver. 2003. Neural mechanisms of general fluid intelligence. *Nature Neuroscience* 6: 316–322.

- Harman, G. 1986. *Change in view: Principles of reasoning*. Cambridge, MA: MIT Press.
- Inzlicht, M., B. D. Bartholow, and J. B. Hirsh. 2015. Emotional foundations of cognitive control. *Trends in Cognitive Sciences* 19: 126–32.
- Killingsworth, M. A., and D. T. Gilbert. 2010. A wandering mind is an unhappy mind. *Science* 330: 932.
- Klee, P. 2012. Art as abstraction. In *Paul Klee*, ed. Eric Shanes. New York: Parkstone Press.
- Koriat, A., and M. Goldsmith. 1996. Memory metaphors and the laboratory/real-life controversy: Correspondence versus storehouse views of memory. *Behavior and Brain Sciences* 19: 167–88.
- Lao Tze (Laozi). 1904. *The book of the simple way*. Trans. W. G. Old. London: Philip Wellby.
- Lebedev, A. V., M. Lovden, G. Rosenthal, A. Feilding, D. J. Nutt, and R. L. Carhart-Harris. 2015. Finding the self by losing the self: Neural correlates of ego-dissolution under psilocybin. *Human Brain Mapping* 36: 3137–3153.
- Leventhal, A. M., R. L. Martin, R. W. Seals, E. Tapia, and L. P. Rehm. 2007. Investigating the dynamics of affect: Psychological mechanisms of affective habituation to pleasurable stimuli. *Motivation and Emotion* 31: 145–57.
- Lindquist, K. A., T. D. Wager, H. Kober, E. Bliss-Moreau, and L. Feldman Barrett. 2012. The brain basis of emotion: A meta-analytic review. *Behavioral and Brain Sciences* 35: 121–202.
- Llinás, R., and S. Roy. 2009. The “prediction imperative” as the basis for self-awareness. *Philosophical Transactions of the Royal Society B* 364, 1301–1307.
- Loftus, E. F. 2003. Our changeable memories: Legal and practical implications. *Nature Reviews: Neuroscience* 4: 231–34.
- Lopez, Jr., D. S. 1992. Memories of the Buddha. In *In the mirror of memory: Mindfulness and remembrance in Indian and Tibetan Buddhism*, ed. J. Gyatso, 21–46. Albany, NY: State University of New York Press.
- Lyubomirsky, S., K. M. Sheldon, and D. Schkade. 2005. Pursuing happiness: The architecture of sustainable change. *Review of General Psychology* 9: 111–131.
- Ma, W. J., M. Husain, and P. M. Bays. 2014. Changing concepts of working memory. *Nature Neuroscience* 17: 347–56.
- Mach, E. 1886. *Contributions to the analysis of the sensations*. New York: Open Court.
- Matthews, L. J., and P. M. Butler. 2011. Novelty-seeking DRD4 polymorphisms are associated with human migration distance out-of-Africa after controlling for neutral population gene structure. *American Journal of Physical Anthropology* 145: 382–89.
- Mauguière, F., and S. Corkin. 2015. H.M. never again! An analysis of H.M.’s epilepsy and treatment. *Revue Neurologique* 171: 273–81.
- Merker, B. 2004. Cortex, countercurrent context, and dimensional integration of lifetime memory. *Cortex* 40: 559–76.

- Michaelian, K. 2011. The epistemology of forgetting. *Erkenntnis* 74: 399–424.
- Momeyer, R. W. 1988. *Confronting death*. Bloomington, IN: Indiana University Press.
- Neisser, U., and R. Fivush, eds. 1994. *The remembering self: Construction and accuracy in the self narrative*. Cambridge: Cambridge University Press.
- Nesse, R. M. 2004. Natural selection and the elusiveness of happiness. *Philosophical Transactions of the Royal Society B* 359: 1333–1348.
- Overall, C. 2003. *Aging, death, and human longevity*. Berkeley, CA: University of California Press.
- Parfit, D. 1971. Personal identity. *The Philosophical Review* 80: 3–27.
- Parfit, D. 1982. Personal identity and rationality. *Synthese* 53: 227–41.
- Patihis, L. 2015. Individual differences and correlates of highly superior autobiographical memory. *Memory* 24: 961–78.
- Perry, J., ed. 2008a. *Personal identity*. Berkeley, CA: University of California Press.
- Perry, J. 2008b. Personal identity, memory, and the problem of circularity. In *Personal Identity*, ed. J. Perry, 135–55. Berkeley, CA: University of California Press.
- Powell, J. W. 1875/2003. *The exploration of the Colorado River and its canyons*. New York: Penguin.
- Putnam, H. 1982. Why there isn't a ready-made world. *Synthese* 51: 141–67.
- Rao, R. P., and D. H. Ballard. 1999. Predictive coding in the visual cortex: A functional interpretation of some extra-classical receptive-field effects. *Nature Neuroscience* 2: 79–87.
- Richards, B. A., and P. W. Frankland. 2017. The persistence and transience of memory. *Neuron* 94: 1071–1084.
- Roseman, I. J. 2008. Motivations and emotivations: Approach, avoidance, and other tendencies in motivated and emotional behavior. In *Handbook of approach and avoidance motivation*, ed. A. J. Elliot, 343–66. New York: Psychology Press.
- Rudd, M., K. D. Vohs, and J. Aaker. 2012. Awe expands people's perception of time, alters decision making, and enhances well-being. *Psychological Science* 23: 1130–1136.
- Ryan, R. M., and E. L. Deci. 2001. On happiness and human potentials: A review of research on hedonic and eudaimonic well-being. *Annual Review of Psychology* 52: 141–66.
- Saldarriaga, L. M., W. M. Bukowski, and C. Greco. 2015. Friendship and happiness: A bidirectional dynamic process. In *Friendship and happiness*, ed. M. Demir, 59–78. Dordrecht: Springer.
- Scharfstein, B. 1998. *A comparative history of world philosophy: From the Upanishads to Kant*. Albany, NY: SUNY Press.

- Schmidhuber, J., 2009. Simple algorithmic theory of subjective beauty, novelty, surprise, interestingness, attention, curiosity, creativity, art, science, music, jokes. *Journal of SICE* 48: 21–32.
- Shoemaker, S. S. 1959. Personal identity and memory. *Journal of Philosophy* 56: 868–82.
- Shoemaker, S. S. 1970. Persons and their pasts. *American Philosophical Quarterly* 7: 269–85.
- Siderits, M. 2007. *Buddhism as philosophy*. Indianapolis, IN: Hackett.
- Singh, S., R. L. Lewis, A. G. Barto, and J. Sorg. 2010. Intrinsically motivated reinforcement learning: An evolutionary perspective. *IEEE Transactions on Autonomous Mental Development* 2: 70–82.
- Still, S., D. A. Sivak, A. J. Bell, and G. E. Crooks. 2012. Thermodynamics of prediction. *Physical Review Letters* 109: 120604.
- Strachan, J., ed. 2003. *A Routledge literary sourcebook on the poems of John Keats*. Oxford: Routledge, 2003.
- Tannenbaum, G., Y. Yeshurun, and S. Edelman. 2009. Trade-off between capacity and generalization in a model of memory. In *30th Annual Conference of the Cognitive Science Society*, 391–96. Washington, DC: Curran Associates.
- Thomson, I., and J. Bodington. 2014. Against immortality: Why death is better than the alternative. In *Intelligence unbound: The future of uploaded and machine minds*, ed. R. Blackford and D. Broderick, 248–62. Chichester, West Sussex: Wiley-Blackwell.
- Tolkien, J. R. R. 1977. *The Silmarillion*. London: George Allen & Unwin.
- Van Boven, L., and T. Gilovich. 2003. To do or to have: That is the question. *Journal of Personality and Social Psychology* 85: 1193–1202.
- Veenhoven, R. 2003. Arts of living. *Journal of Happiness Studies* 4: 373–84.
- Vidal, F. 2016. Desire, indefinite lifespan, and transgenerational brains in literature and film. *Theory and Psychology* 26: 665–80.
- Wacongne, C., E. Labyt, V. van Wassenhove, T. Bekinschtein, L. Naccache, and S. Dehaene. 2011. Evidence for a hierarchy of predictions and prediction errors in human cortex. *Proceedings of the National Academy of Science* 108: 20754–20759.
- Walker, M. 2014. Uploading and personal identity. In *Intelligence unbound: The future of uploaded and machine minds*, ed. R. Blackford and D. Broderick, 161–77. Chichester, West Sussex: Wiley-Blackwell.
- Wickelgren, I. 1997. Working memory linked to intelligence. *Science* 275: 1581.
- Williams, B. 1973. The Makropulos case: Reflections on the tedium of immortality. In B. Williams, *Problems of the self*, 82–100. Cambridge: Cambridge University Press.
- Winkielman, P., K. Berridge, and S. Sher. 2011. Emotion, consciousness, and social behavior. In *The Oxford Handbook of Social Neuroscience*, ed. J. Decety and J. T. Cacioppo, 195–211. Oxford: Oxford University Press.