

PERSONAL VOLATILITY

Meghan Sullivan*

Last year, I faced a decision. I'd saved up \$3,000. I could put it towards my retirement account, or I could spend it to attend a friend's wedding in India. I was lucky to even have the decision—the vast majority of Americans do not have access to any employer-sponsored retirement plan and face crippling near-term expenses that make personal saving impossible.¹ I grew up in a family that lived paycheck-to-paycheck, and to this day my parents have little by way of savings. To some extent, my adult life has been defined by trying to overcome our family's anxiety about money and do a better job of planning for the future.

Which is to say, I try hard to base my savings preferences and activities on reasons. One hazard of being a philosopher is that I'm trained to seek out reasons... the pros and cons can quickly multiply.

I'm 34 right now. If I put the money in the retirement account, I won't be able to access it without penalties until I am 65.

Pros: I am pretty confident that the index funds in my retirement account are a sound vehicle for saving money. If the "rule of 72" and broad historical trends in the market hold fast—and I think they will—then \$3,000 invested now will likely be worth almost \$24,000 by the time I withdraw it. Even adjusting for inflation, there is no way I'd spend that much money now on a trip to India. So maybe I should save.

Cons: that happy withdrawal date is another lifetime from now! I live a (mostly) healthy lifestyle. I am confident I will survive that long. But I also think I will be quite a different person by the time I reach 65. I will still like money, but my priorities will change. And I'm not presently very excited at the prospect of waiting 31 years in order to blow \$24,000 on a downpayment for a seniors-only condo in Miami or a tranquil cruise down the Danube River. While I think I will highly

*Thanks to audiences at Columbia University, the 2017 Desert Philosophy Workshop, and the 2017 Pacific APA Transformative Experience Pre-Conference.

¹See for instance Kim (2015) and Gabler (2016).

value the condo or cruise *then*, I'd much prefer to spend the \$3,000 *now* on the wedding trip. After all... it's India.

In the end, the prospect of adventure on the subcontinent was too much for me. I booked the trip. Was my decision rational, in the sense that it was properly supported by reasons? Or was I falling into the same spending trap that has been my parents' financial undoing?

1 Two Kinds of Volatility

Prudentially rational agents pay attention to the potential for change. For instance, suppose I am comparing potential retirement investments in the US bond market and Greek bond market. I should be a lot less confident that I know what will happen to money put in the Greek bonds since their economy is more likely to change in unpredictable ways. The Greeks need to sweeten the deal for potential investors, selling their bonds at a low price now to compensate for the risk. Let's call the range of probabilities assigned to potential outcomes of a choice the choice's *outcome volatility*. The wider the range of potential outcomes and probabilities, the more volatile the choice. And when you discount the present value of an option based on its probability of occurring, we can call this *probabilistic discounting*. Rationality requires you to take outcome volatility into account, though of course the probabilities are not always easy to judge. Much work in decision theory aims to offer an account of how rational agents should act given their approximations of outcome volatility.

But what if you also think that *you* are volatile over long intervals of time? How should assumptions about how you will change figure into your planning for the future? Let's call the range of your variation over time your *personal volatility*. For example, for most normally developing adolescents, the interval from ages 10-19 is highly personally volatile. The shorter interval from age 19 to 20 is significantly less so. When you discount the present value of an option because of changes you predict you will undergo before the scheduled outcome is realized, we can call this *personal discounting*.

Personal discounting and probabilistic discounting can come apart, as my India splurge illustrates. I am confident enough in the probabilities on each of my two options (spend or save). I just don't now care much about myself so far out, since I predict that I will change significantly in the interim. This indicates another key difference between the forms of volatility. Outcome volatility

is measured using the probabilities of different future events occurring, where those events can be described aperspectivally (i.e. the problem can be posed as there is an 80% chance that Sullivan will withdraw \$24,000 in 2040 and a 20% chance she will not). But the question of personal volatility only arises from the first-personal, temporally-situated point of view (i.e. does Sullivan in 2040 matter *to me, now*, in a self-interested way.)

Even raising the question of personal discounting requires wading into the debate about the metaphysical and psychological indispensability of subjectivity. *Whole-hearted reductionists* about personal identity believe (1) that the facts about a person's identity over time just consist in the holding of certain more particular facts that can be described entirely without reference to a subject, and (2) that these impersonal facts are the only facts that are relevant to rational planning.² In contrast, *non-reductionists* deny both (1) and (2). And *half-hearted reductionists* think subjectivity could be eliminated from their complete metaphysical description of the world, but they think subjectivity is an ineliminable part of rational planning.³

Whole-hearted reductionists have no special concern for personal volatility. If Derek Parfit is taken at his word, he was once a whole-hearted reductionist. "I find (reductionism) liberating, and consoling. When I believed that my existence was a further fact, I seemed imprisoned in myself. My life seemed like a glass tunnel, through which I was moving faster every year, and at the end of which there was darkness. When I changed my view, the walls of my glass tunnel disappeared."⁴ But few of us are capable of sustaining this attitude, at least psychologically. Parfit himself eventually lost heart and denied whole-hearted reductionism.⁵

Nor is it obvious what it would be to adopt such a detached view of our lives. Even if sometimes you can imaginatively "step outside" yourself and consider your life as a connected stream of stages, the bird's eye perspective on your life is still *a perspective*, imagined from it's own apparently personal, present vantage. Much of the literature on personal identity presupposes that we have secure intuitions about what rationally matters to us (or would matter in a counterfactual scenario). Popular theories use these intuitions as evidence for various theories of what it is to be a persisting self or, at least, what we take ourselves to be. For instance, David Velleman sets out to argue that

²For instance, Parfit (1984).

³For instance, Lewis (1979) or Paul (2014). For more on this as a metaphysical view, see Section 3 of Zimmerman (2005).

⁴Parfit (1984)[281].

⁵Parfit (1999).

we care most about occupying selves with perspectives (rather than strict numerical persistence). He argues for this using data about concern: “What we most want to know about our survival... is how much of the future we are in a position to anticipate experiencing. We peer up the stream of consciousness, so to speak, and wonder how far up there is still a stream to see... My question is how long there will be someone to occupy the position that is the center of my self-centered projections.”⁶ On a different track, Jennifer Whiting thinks our patterns of actual or counterfactual rational egoistic concern give evidence that we are most concerned with which future or past selves we can bear an analogue of the friendship relationship toward.⁷ David Lewis thinks that because both survival and psychological connectedness matter for rational egoistic concern, we should think of ourselves as four-dimensional perdurants.⁸

I confess, I find this direction of argument puzzling. Velleman and Whiting, for instance, ask to imagine reality as though it is a field of candidate selves, and we need to figure out which “selves” we’d care to occupy or befriend. But I already had a very strong pre-philosophical understanding of what it is to be myself, even that understanding eludes metaphysical specification. I believe I’ve persisted for almost 34 years, that I have many properties which now give me reason to plan for the future in certain ways, and that I also have some biases which sometimes prevent me from properly attending to my reasons. My more pressing philosophical concern is to figure out what I should care about. I am much more likely to revise any of my views about rational egoistic concern than I am to revise my self-conception.

This essay is aimed at the half-hearted reductionists and non-reductionists. If you acknowledge a first-personal perspective, either in reality or as an ineliminable part of your psychology and planning, then is personal discounting ever rational? Or is personal discounting just one of our many psychological biases that prevent us from living prudently—like our inability to properly estimate future values and our inability to control our strongly time-biased emotions? Much of the literature on personal identity focuses on questions like what is it to be a persisting self, asking us to imagine changes and form judgments about whether we’d survive or still exhibit self-concern after those changes. Instead, I am going to assume that you already know (more or less) that you are some kind of persisting self capable of enduring somehow through change. I will ask what patterns of

⁶Velleman (1996, 67-68).

⁷Whiting (1986).

⁸Lewis (1976).

self-concern you ought to have, given what you think you are.

I will consider two strategies for defending personal discounting over long intervals of time. The first strategy focuses on qualitative personal volatility. Most of us will change significantly with respect to *how* we are over long intervals. Our personalities, quasi-memories, and aspirations will change.⁹ Our bodies will change with respect to our physical abilities, our features, and many of our parts. If similarity of qualitative properties is what matters for self-interest, then qualitative volatility would be a good basis for personal discounting.

The second strategy focuses on numerical personal volatility. Over long intervals, are we likely to change with respect to *how much* we are? And would this be a ground for thinking we are less of ourselves after change? If self-interest depends on facts about persistence and you are less-and-less you in the more distant future, then such numerical volatility would be a good basis for personal discounting. I will also consider whether there are possible metaphysical changes (i.e. fission) that could provide a rational basis for personal *mark-ups*—increases in future directed egoistic concern.

I'll try to convince you that, despite their initial appearances and auspicious philosophical lineages, these two strategies for defending personal discounting fail. The philosophical literature on egoistic concern is typically dominated by thought experiments involving teletransportation, brain dissection and amnesia. But I'll pursue a different methodology, drawing on thought experiments from more realistic cases of volatility and empirical work on egoistic concern. The stakes are significant. Most of us are not doing nearly enough to plan for our distant future, and it is not usually because we have grim predictions about what we are, how long we will live, or what will happen to the world economy. More likely, it is because we are in the grip of biases in our egoistic concern.

2 A Qualitative Basis for Egoistic Concern?

Singapore is a nation obsessed with planning. NTUC Income is a large investment firm on the island, and for the past few years they have been running a campaign called “Future You”. At bus stops, on billboards and in movie theaters, they have posters and videos depicting a young professional (roughly my age) being greeted by his/her self from the future. In some of the ads, the future self warmly embraces the startled younger self. In others, the future self beats the younger

⁹A quasi-memory is a memory-like mental state that does not presuppose facts about persistence through time.

counterpart. Each ad ends with “How will the Future You thank you?”¹⁰ The implied message: open a retirement account with NTUC or fear the wrath of your future self.

It is no surprise that NTUC wants you to start an investment account—they’ll make commissions right now on every stock you trade. Likewise, it is no surprise that the United States government has a vested interest in my making contributions to a retirement scheme—every contribution I make takes a bit of pressure off of the Social Security system. But we don’t just think saving is good because it benefits banks or governments. (Very little of *my* altruism is directed at bankers.) The implied message of the NTUC ads is that it is in your *self* interest to make these kinds of sacrifices. You are meant to see the ads and think: “That future person is just as much *me*. I should care about her as much as I care about myself now.”

But what is it about that pensioner in the Future You ad that should make him/her matter to you in a self-interested way, rather than, say, in the more broadly moral way we care that the elderly have adequate financial resources? (We will assume for the foregoing arguments that there is a difference between caring self-interestedly and caring morally.) One flat-footed answer is that you just care that the future person will be *identical to you now*. But there are some puzzles for thinking mere identity is the basis of egoistic concern. Consider an example from Thomas Nagel. Suppose it turns out that people just are their bodies. When you die, you will permanently become a corpse. Would the discovery that you’ll continue to exist as a corpse affect your rational planning? Probably not. That hunk of flesh embalmed in formaldehyde and buried in the ground some hundred years from now might be you, but you will care a lot less (if at all) about yourself *then*. Barring special religious reasons or moral concern for those we leave behind, most of us are unconcerned for our future corpses.¹¹ Thought experiments like this lend support to the view that our grounds for egoistic concern are something more substantive than just mere facts about identity.

A variety of candidates have been advanced as the “missing ingredient” that we care about when we care self-interestedly. One of the most prominent views (suggested by John Locke and advanced especially by Derek Parfit) is that the basis for egoistic concern is psychological connectedness. Indeed, Parfit famously holds that strict identity doesn’t matter at all, once we realize that connectedness is what matters.¹² Direct psychological connections come in two varieties. One variety

¹⁰To see the website for the ad campaign: <http://futuremadedifferent.com>.

¹¹Nagel (1970).

¹²“Connectedness is a more important element in survival... Even if connectedness is not more important than

of connection is similarity: you share the same memories, personality traits, preferences, and beliefs. Another variety of connection is causal: your choices at a time cause your future memories, and your intentions at a time cause your future satisfaction or frustration. For most normally developing individuals, psychological connectedness comes in degrees. I am much more connected with the woman I was a year ago or will be a year hence than I am with my teenage self (who hated philosophy) or with my elder self (who will love slow river cruises). Psychological connectedness is also an intransitive relation, which explains why we might exhibit personal volatility over long intervals even if every stage of our life is strongly directly connected to her immediate stages.

Within psychological connectedness theories there is also significant debate about whether the basis of egoistic concern must hold one-one or could hold one-many. For instance, if you were scheduled to undergo fission in the near future, with the process resulting in two distinct selves with psychologies very strongly connected to yours now, would you care self-interestedly and equally about the fates of each of your daughter selves? Or are we only able to care self-interestedly about at most one future self at a time? And which of your duplicates should get access to your retirement fund? Parfit denies that strict identity is a basis of egoistic concern because he takes seriously the possibility of this type of numerical personal volatility. We will return to this issue in Section 6.

3 The Concern Argument for Personal Discounting

All but the most boring of us exhibit qualitative personal volatility. Our apparent memories fade, our personalities morph, our projects and intentions change directions. This suggests a general argument for why personal discounting is typically permissible over long intervals, one based on the nature of egoistic concern:

The Concern Argument for Personal Discounting

- (1) There is some criterion C (like psychological connectedness) that determines your egoistic concern, and it is intransitive and comes in degrees.
- (2) It is prudentially rational to discount events based on the degree to which those events have the basis for your egoistic concern.

continuity, the fact that one of these is a relation of degree is enough to show what matters in survival can have degrees." Parfit (1971, 21). Compare to Lewis, who thinks identity and connectedness both matter. Lewis (1976).

(3) Given that you are a normally developing agent, as events are scheduled in the far future, C will tend to hold to a lesser-and-lesser degree. (The Qualitative Volatility Assumption)

C. As events are scheduled in the far the future, it is rational to discount them.

Note the argument doesn't work if psychological continuity—a transitive relation—is the basis of egoistic concern. You are psychologically continuous with any future self that is connected to your present self via a series of unbroken relations of psychological connectedness. Premise (2) seems to follow from the assumption that prudential rationality is a matter of self-interest—if an event has less of whatever matters to self-interest, it is perfectly rational to care less about it. But, as stated, the premise is ambiguous between what actually describes your pattern of egoistic concern and what you take to be a reason-giving basis for your egoistic concern. Premise (2) is only true on the second reading. So we should make the argument more precise by clarifying the first two premises:

The Concern Argument for Personal Discounting—Revised

(1*) There is some criterion C (like psychological connectedness) that determines your *rational* egoistic concern, and it is intransitive and comes in degrees.

(2*) It is prudentially rational to discount events based on the degree to which those events have C.

(3) Given that you are a normally developing agent, as events are scheduled in the far future, C will tend to hold to a lesser-and-lesser degree. (The Qualitative Volatility Assumption)

C. As events are scheduled in the far the future, it is prudentially rational to discount them.

The Concern Argument, as stated, is highly general. It accommodates the prevailing philosophical view that self-interest is a matter of psychological connectedness. But it could equally work for some other theories of egoistic concern. For instance, if you think preservation of your body or your brain is what you really care about, then a bodily-connectedness or neuro-connectedness criterion of egoistic concern could be plugged into the argument. Jeff McMahan offers a version of

the Concern Argument to justify caring less about yourself in the more distant future, and he thinks the basis of egoistic concern is a combination of psychological and neurological factors. According to McMahan, we minimally care about having our consciousness connected over time (and he assumes consciousness is realized in the brain). After we've satisfied ourselves that some future self's consciousness will be the same as ours, *then* we look to see how psychologically connected they are to us. We care more or less self-interestedly about that future self to the extent that they are more or less psychologically connected with us.¹³

4 Simple Connectedness and Moral Valence

The Concern Argument might seem initially appealing, but it is unsound. In this section, I will offer arguments against premise (1*), at least insofar as it purports to represent our actual judgments about egoistic concern. And I will argue that attempts to make premise (1*) psychologically realistic will undermine premise (2*). We'll use as our case study the hypothesis that psychological connectedness is the basis for discounting. Then I will generalize the objections to other candidates for C.

Premise (1*) of the Concern Argument can be understood in two ways. First, it might be thought to make both a psychological prediction and a normative judgment. That is, C predicts our pattern of egoistic concern and that pattern is rational. Call this the *psychological reading*. Second, it might be thought to be only a revisionary normative judgment. That is, you might not in fact care at all about C but you rationally should. Call this the *revolutionary reading*. The Concern Argument looks significantly more plausible if premise (1*) is given the first reading. On the revolutionary reading, we need (at least) an independent argument for why C matters as well as a theory of error for why you are ignorant of your self-interest.¹⁴

On the psychological reading, premise (1) makes (in part) a prediction which is empirically testable, namely:

The Simple Connectedness Hypothesis: We will have diminished egoistic concern for any selves who are psychologically dissimilar or causally disconnected from our present self.

¹³McMahan (2002, 79-80).

¹⁴Thanks to Johann Frick and Trenton Merricks for discussion here.

But there are problem cases for this as a hypothesis about our actual patterns of concern. Consider the true story of Susannah Cahalan’s brain disorder:

Brain on Fire:

In 2009, Susannah Cahalan had a sudden, radical mental breakdown. Until that point she was a healthy 24-year old living in New York and reporting for a major newspaper. She awoke one morning in the midst of a grand mal seizure. After that event, she lost her memory, experienced extreme paranoia and psychosis, and soon after lost most of her emotional responses. After a month of living in this radically debilitating state, a doctor successfully diagnosed Cahalan with a rare form of encephalitis. The key to the diagnosis came when Cahalan was asked to draw and number a clock face. She drew a normal circle, but bunched all of the hours to the right-hand side—a common symptom of inflammation in the right side of the brain. After a course of treatment for encephalitis, Cahalan mostly returned to her pre-disease psychological functioning. She remembers life before the disease, but had to piece together life during her “month of madness” by watching hospital security footage and interviewing her caretakers.¹⁵

During her illness, Cahalan was not able to form new memories or preserve the old dimensions of her personality. The Simple Connectedness Hypothesis predicts that Cahalan should feel like a “new woman” after her recovery, not identifying in any significant way with her hospitalized self and troubled by the long disruption in the psychological connections with her previous self. It also predicts that she should have significantly diminished concern for what happened in her life during the period of the disease. In fact, neither of these hypotheses are true. Cahalan sees herself as having *recovered* from her disease—carrying on a life that now includes this horrific adventure as an episode. And she cared deeply enough about her life during the disease to watch tapes and conduct interviews, trying to understand what it was like for her when she was ill.

Imagine yourself in Cahalan’s predicament before her disease took hold. But instead of being surprised, imagine you know you’ll abruptly lose your psychological faculties for a month. Would this information make it less rational for you to plan for your future during the disease? I suspect many of us would make substantial sacrifices now to ensure we were well cared-for during such a disease.

But perhaps not all of us would be so concerned to recover our encephalitic selves. Is this evidence that psychological connectedness is the basis of egoistic concern? Parfit, McMahan and other proponents of the psychological connectedness accounts of egoist concern typically motivate

¹⁵Cahalan (2013).

the criterion by asking you to imagine a radical change, then judge how interested you'd be in your life after the change, and reflect on whether this concern is reasonable. The examples almost always involve a change with a negative moral or social valence. Either you lose your autonomy during the change (i.e. you contract amnesia or you are to be restrained and tortured). Or you change in such a way as to join a class that is typically accorded less social respect (i.e. you become elderly or you become cognitively disabled). Or you change in such a way as to join a class that many readers would find morally repugnant (i.e. you transform from a generous socialist to a greedy oligarch or you transform from a human into a vampire).¹⁶

But the Simple Connectedness hypothesis predicts that our egoistic concern will also diminish when the psychological disruptions are neutral or moral/social improvements. Is this right? Consider a case with a disruption analogous to an encephalitic episode but without the element of severe disability. Canadian pianist Glenn Gould was famous for his ability to enter “flow” states while performing. Reporter Robert Krulwich reports the experience of Gould playing Bach’s Partita #2:

Flow:

“Glenn Gould, supreme interpreter of Bach, is sitting at his living room piano on a low, low chair, his nose close to the keys. He’s at his Canadian country house in his bathrobe. Through the window, you catch snatches of his back yard. It’s a windy day and he’s got a coffee cup sitting on the piano top. He’s working on a Bach partita, not just playing it, but singing along in his swinging baritone. As he plays, he gets so totally, totally lost in the music that suddenly, smack in the middle of a passage, with no warning, for no apparent reason, his left hand flips up, touches his head; he stands up, and walks in what looks like a trance to the window. There’s an eerie silence. Then, in the quiet, you hear the Bach leaking out of him. He’s still playing it, but in his head, he’s scattling the beats. Then he turns, wanders back, sits down, and his fingers pick up right where his voice left off, but now with new energy, like he’s found a switch and switched it.”¹⁷

During his “rapture” Gould is psychologically disconnected from his previous selves. He’s so absorbed in the music that he is unaware of his surroundings. He cannot call to mind memories. His personality is significantly altered. He is completely disinterested in the future. And his actions lack any conscious intention. “Flow” states like this are the subject of significant research in positive psychology.¹⁸ Skilled surgeons report entering such states during long procedures; in one case a

¹⁶Parfit (1984, 327) and Chapter 1 of Paul (2014).

¹⁷Krulwich (2014).

¹⁸Nakamura and Csikszentmihaly (2002).

surgeon was so absorbed as to be unaware that the operating room ceiling had collapsed. Athletes report entering such a state during intense competition.

Presumably entering “flow” does not diminish egoistic concern. For instance, a surgeon wouldn’t take less credit for a complicated procedure performed while in this disconnected state. Knowing that you will flow during your marathon does not make the experience less important to you now. But the Simple Connectedness hypothesis predicts that flow would diminish egoistic concern.

You might think that the duration of the gap makes a difference. So consider a case where there is a longer psychological disconnect, analogous to Cahalan’s break.

Meditative Retreat:

Jonathan has booked a month-long meditation retreat in Sedona, Arizona. In his civilian life, Jonathan is an extroverted and stressed-out lawyer. But during the retreat, he will enter a month-long state of flow. He’ll be acutely aware of his breathing and bodily movements. He will think of neither the future nor the past. He won’t have goals or intentions. His mind will be “quiet”. He knows he will become relaxed and pacific, at least for the duration of the retreat.

Should Jonathan now be less concerned for himself during the retreat? Most of us, I suspect, think not. Indeed, after the retreat, when Jonathan gets back to his stressed, extroverted default, it would be normal for him to identify just as strongly with his calm, pensive retreat self.

You might also think the encephalitic and flow cases differ because in the latter cases the changes are intended but in the former case they are not. Intention is one form of direct psychological connection. McMahan appeals to intention connections to explain why self-improvement would be rational on his version of the psychological connectedness theory.¹⁹ But we can also imagine the flow cases in a way where Gould or Jonathan just “slip” into these states of absorption and tranquility. Unintentional flow does not seem significantly more disruptive to egoistic concern than intentional flow.

A more salient difference between the encephalitis case and the flow cases concerns the culturally-mediated moral worth we assign to the agents in the episodes of radical change. In the encephalitis case, Cahalan is an object of pity. She’s seriously disabled, dependent upon her caregivers, and emotionally hostile. In the flow cases, Gould is demonstrating virtuosity. Jonathan is becoming his calmer, more introspective and “better” self. Episodes of psychological disconnect that are morally

¹⁹See section 1.5.2 of McMahan (2002).

neutral or moral improvements do not seem to diminish egoistic concern. This suggests another hypothesis for the basis of egoistic concern, which we might call:

The Moral Valence Hypothesis: We will have diminished egoistic concern for any selves who are psychologically dissimilar or causally disconnected from our present self only if those selves have perceived diminished moral capacity or diminished social/moral worth, where such worth might be mediated by cultural assumptions.

There is direct empirical evidence of the Moral Valence Hypothesis in the so-called “Phineas Gage” effect in psychology. Phineas Gage was a railway worker who in 1848 survived an accident where an iron spike was driven through his frontal lobe, permanently and radically altering his personality. His case became a focal point for work on the biological underpinnings of personality and emotion. Subjects in a series of Gage-inspired personality psychology studies are asked to imagine one of two scenarios. In the first, a protagonist (named Phineas) suffers a brain injury (and corresponding psychological disruption) and as a result becomes significantly crueler. In these cases, subjects are less likely to report that Phineas is the same individual after the ordeal. But if instead asked to imagine Phineas becoming significantly *kinder* after his brain injury, subjects are more likely to report that he is the same individual.²⁰ Third-party judgments of personal persistence are sensitive to the moral valence of the psychological changes.

Other studies suggest that we have a robust moral valence when it comes to attributing responsibility and agency, sometimes called the “Pollyanna Principle” after the relentlessly optimistic character in Eleanor Porter’s novel *Pollyanna*. George Newman, Paul Bloom and Joshua Knobe conducted a series of experiments measuring our attitudes about whether some behaviors are manifestations of an agent’s “true self” rather than a reflection of some external feature of the agent’s situation. Subjects are more likely to attribute responsibility to an agent’s “true self” when the behavior is perceived as morally valuable. And they are more likely to assign a diminished role to an agent’s “true self” when the behavior is perceived as immoral.²¹ This study, among others, lends support to the view that we have a positivity bias in our judgments about personal persistence—namely, we tend to think true selves are fundamentally moral. This positivity bias offers further evidence for the Moral Valence hypothesis. We are less able to see our “true selves” present in an

²⁰Tobia (2015).

²¹See Newman et al. (2014).

imagined scenario where psychological volatility has perceived negative moral consequences.²²

Note that these effects hold whether the qualitative personality changes are intentional or accidental, temporary (as in the flow cases) or permanent (as in the Gage cases).

5 Egoistic Concern and Persistence Bias

With these effects in mind, let's return to the revised Concern Argument:

(1*) There is some criterion C (like psychological connectedness) that determines your rational egoistic concern, and it is intransitive and comes in degrees.

(2*) It is prudentially rational to discount events based on the degree to which those events have C.

(3) Given that you are a normally developing agent, as events are scheduled in the far future, C will tend to hold to a lesser-and-lesser degree. (The Qualitative Volatility Assumption)

C. As events are scheduled in the far the future, it is prudentially rational to discount them.

In the previous section, I offered support for the Moral Valence hypothesis. Briefly: diminished psychological connections leads to diminished egoistic concern *only* when the change is also accompanied by a loss of moral or social standing. And unqualified psychological connectedness does not accurately predict our patterns of egoistic concern. This is evidence against premise (1*) of the Concern Argument, at least insofar as simple psychological connectedness is put forward as the candidate for C. There are many cases where you are unlikely to judge diminished psychological connection as a reasonable basis for discounting.

It is also evidence against (2*). Our positivity bias is a *bias*. It is a valuing attitude that is sensitive to features of the situation which are not reason-giving.²³ Premise (2*) of the Concern

²²Korsgaard hints at a responsibility-focused hypothesis like this in her criticism of Parfit, Williams and the usual cases invoked in the debates about egoistic concern: "These writers usually emphasize the facts that after the surgical intervention we are altered, we have changed. But surely part of what creates the sense of lost identity is that the person is changed by *intervention*, from outside. The stories might affect us differently if we imagined changes initiated by the person herself, as a result of her own choice. Authorial psychological connectedness is consistent with drastic changes, provided those changes are the result of actions by the person herself or reactions for which she is responsible." Section IV of Korsgaard (1989).

²³For more on this account of bias, see Chapter 3 of Sullivan (under contract).

Argument is only plausible if you think the qualitative basis for your egoistic concern is reasonable. But once we realize our judgments about egoistic concern exhibit this moral valence, we shouldn't think these judgments give us good reasons to discount. It simply isn't the case that we are only authors of the morally valuable chapters of our lives.²⁴ Rationality requires a certain threshold of non-arbitrariness in our reasons. So it requires that adopt a unified attitude toward the cases. Either you should care just as much about your encephalitic self or such intensive yoga retreats are never in your self-interest. I submit that in response to the tension, we are more likely to seek a less demanding (and less volatile) basis for egoistic concern.

Are these problems specific to the psychological connectedness theory? We can likewise imagine someone planning around a radical change in her body (if bodily connectedness is the basis of concern) or planning around radical restructuring of her brain (if neuro-connectedness is the basis of concern). There is less empirical work on how our attitudes about bodily or neural volatility affect our egoistic concern, but there is some reason to expect they exhibit the same phenomenon of moral valence. Egoistic concern is likely to diminish if you imagine your body will be radically altered by amputation. But egoistic concern will remain stable if you imagine your body altered by radical weight-loss. Egoistic concern is likely to diminish if you imagine suffering brain damage that makes you somewhat crueler or less sensitive. But it is likely to remain stable if you imagine that your intensive yoga retreats will result in neural re-wiring. These hypotheses are testable, just as the hypotheses about the psychological basis of egoistic concern are. They demonstrate, at the very least, that there is insufficient evidence that a qualitative criterion for concern can be found which satisfies both (1*) and (2*) of the Concern Argument.

This leads to a much more general methodological worry with the Concern Argument. Philosophers have a nasty habit of telling people what they care about. Using thought experiments that might double as science-fiction plots, Parfit, McMahan, Velleman and others try to motivate one or another theory of the basis of rational egoistic concern, at the same time using these intuitions to offer evidence about what matters to us vis-a-vis being a self over time. If you enter a tele-transporter that copies your mind into a new body, would you care if the old body were destroyed? No? You have a psychological connectedness view of selfhood. If your brain were dissected at the

²⁴Unless you have a radically voluntarist conception of personal persistence, such that you think it is possible to metaphysically omit yourself from events with a negative moral or social valence.

corpus callosum and each hemisphere were transplanted in a new body, which of the bodies would be entitled to your assets? Would you be willing to give each new brain hemisphere half of your estate? Then you have a neuro-connectedness view. One thing we know about our intuitions in such thought experiments is that they are highly sensitive to how the cases are described.²⁵ Told one way, it might strike you as though you were traveling through spacetime, splitting in two or switching bodies. Told another way, you stay right where you are but undergo a radical change. The shiftiness of these thought experiments means they have never been great evidence for a theory of egoistic concern.

What *is* good evidence for a theory of rational egoistic concern? The empirical studies cited above are a great start, at least insofar as we think our ordinary patterns of egoistic concern are rationally defensible. The studies indicate a bias in our attributions of persistence and patterns of concern. Indeed, to the extent that debates about the metaphysics of personal identity have turned on data about rational egoistic concern, it is astounding how infrequently philosophers rely on empirical data about the conditions under which subjects think they have reason to care about qualitative volatility.

We'd also do well, I think, to take seriously the received wisdom about the rationality of distant future planning. For thousands of years and across cultures, it has been taken as a norm that agents ought to plan for their futures, even when those futures are quite a long ways off.²⁶ This wisdom is enshrined in fables like Aesop's *Ants and the Grasshopper*. It is assumed in the plethora of quasi-academic self-help books recommending you develop grit, practice 10,000 hours, and wait for the second marshmallow.²⁷ It has struck many as a platitude of commonsense that prudent individuals keep the long-run in mind. So whatever candidate properties we advance as being the rational basis of egoistic concern, they must be reconciled with the robust assumption that we ought to pay significant attention to our self-interest over the long run. The Concern Argument for personal discounting faces a stiff challenge from the outset.

So what *should* we care about when we care self-interestedly? One obvious candidate for a qualitative basis of egoistic concern is the mere persistence and connectedness of our consciousness (and, indirectly, whatever is needed to realize our consciousness). This would justify Cahalan's

²⁵See for instance Williams (1970) on the "changing bodies" thought experiments.

²⁶Or in the case of some Far Eastern cultures, agents ought to plan for the futures of their particular families.

²⁷See Duckworth (2016), Gladwell (2008), and Mischel (2014).

reaction in the encephalitis case—she may not have her personality through the ordeal but she knows she was still awake and having experiences. Flow cases, likewise, do not disrupt mere consciousness. The trouble is, as we posit a weaker and more tolerant basis for rational egoistic concern, premise (3) of the Concern Argument looks less and less plausible. While our psychologies might be volatile over long periods of time, our consciousness persists relatively unperturbed.

6 The Persistence Argument for Personal Discounting

So far we've considered two forms of volatility that might make a rational difference to distant future planning. The first is outcome volatility—the range of probabilities assigned to potential outcomes of a choice. It is rational to discount events probabilistically, and as a matter of course we often should deem events less and less probable as they are scheduled further in the future. But this is not enough to explain why we save so little, since most of us are quite optimistic about our survival in the next few decades. The second is qualitative personal volatility—changes in whatever properties are a reasonable basis for egoistic concern. In the previous three sections, I raised doubts that we are qualitatively volatile in any way that would make a difference to rational egoistic concern. We turn now to a third, somewhat more nebulous form of volatility that might justify personal discounting: numerical volatility.

To introduce this form of volatility, consider another true story:

Brain on Ice:

In 2011, Kim Suozzi, age 23, was diagnosed with terminal brain cancer. Suozzi had some training in neuroscience and had been involved in research on connectomes—the neural networks that some scientists believe house an individual's consciousness. Suozzi contracted with a company called Alcor to cryogenically preserve her brain upon death. In January 2013, she slipped into a coma and was pronounced dead. Technicians from Alcor separated Suozzi's head from her torso, pumped her brain with cryoprotectant, and stored it in a container in their facility. The plan required creating a trust of \$80,000 to keep her brain preserved. Suozzi's hope was that at some future time, technology will advance enough that her connectome can be mapped and simulated. She hoped such a procedure might bring her back.²⁸

Suppose in fifty years we will have the technology to map the connectome of a preserved brain. And suppose Suozzi's connectome will be simulated in new brain tissue. Will Suozzi come back? Was

²⁸Harmon (2015).

Suozzi rational to create a trust for such an event before she succumbed to her cancer? You might think it is an all-or-nothing matter whether Suozzi (or anyone) could be revived in this way. If you think existence is an all-or-nothing affair, then there is a determinate fact about whether a revived cryogenic brain is or isn't identical to its pre-mortem donor. In this case, the riskiness of freezing a brain is represented as a form of outcome volatility—it is hard to be confident the procedure will work.

But you may also think that existence over time is not an all-or-nothing affair. The properties that determine whether you remain you over a series of changes can come in degrees. As a result, it can be a matter of degree how you (or anyone) would persist through an adventure like Suozzi's. Suppose that a person's existence over time is determined by her connectome—she survives over time just in case there is a series of connections between her neural network at one time and a neural network at another. Connectome relations can presumably hold in degrees. Between 2010 and 2011—before her disease—Suozzi's connectome changed little. But as her cancer progressed, her brain changed rapidly. The cryogenic freezing process further damaged her connectome. Any neural network built from her brain after 2013 will be related to her 2011 connectome, but tenuously. You might think that means that if Suozzi is revived, the resulting person will be much less Suozzi.

A gradualist view of persistence like this suggests another potential defense of personal discounting:

The Persistence Argument for Personal Discounting

- (1) There is some criterion C that determines your persistence over time, and it is intransitive and comes in degrees.
 - (2) It is prudentially rational to discount a future event based on the degree to which you will be affected by that event.
 - (3) For any plausible candidate for your C, as events are scheduled further in the future, C will tend to hold to a lesser degree. (The Numerical Volatility Assumption)
- C. As events are scheduled further in the future, it is prudentially rational to discount them.

As with the Concern Argument, candidates for C include psychological connectedness, bodily con-

nectedness, or embodied consciousness. But unlike that argument, the Persistence Argument says nothing about what qualities you care about into the future. Instead, it turns on how you rationally care about the “amount” of yourself into the future.

This is a potential route to defending personal discounting. But all three premises of the persistence argument are deeply controversial.

There are two ways we might understand Premise 1. First, you might think that facts about your persistence are susceptible of indeterminacy. When your persistence comes in degrees, this means there is an increase in indeterminacy about your persistence every time C is disrupted. Call this the *indeterminacy* reading. This reading rules out the possibility that your identity is determined by something that cannot come in degrees, like having the same soul over time. Many philosophers—myself included—have the view that existence over time is an all-or-nothing matter. For anything, after any change, there is a determinate fact about whether the thing undergoing the change has survived.²⁹ And it is never indeterminate how many things exist, selves included.³⁰

On the indeterminacy reading, premises (1) and (3) also require a dramatic departure from our ordinary self-conceptions. For instance, few of us believe that our age is radically indeterminate, but if (1) and (3) are true, the parallel reasoning applies equally to our past. As events are scheduled further in the past, it happened less to me. So I should not be so quick to assert that I am 34 years old. At best I was indeterminately present in the 1980s. If we assume psychological connectedness is the relevant C, then when Susannah Cahalan relates stories about her adult life, she should include gaps when she is at best indeterminately the protagonist. I’ve maintained that theories of prudential rationality should be *prima facie* conservative with respect to what they entail about our self-conceptions. That is, any normative argument with a controversial metaphysical premise is likely to be question-begging (or at least deeply unpersuasive) because it is typically much easier to revise our confidence about what is rational to do or prefer than it is to revise our confidence about what we are.

You might also think that facts about existence and persistence are always determinate, but selves are like Big Gulps... they come in different sizes. Suozzi could determinately survive by having her connectome simulated, but there will be less of her after the procedure.

²⁹Sullivan (2012).

³⁰Sider (2001, 120-139).

Premise (2) of the Persistence Argument looks plausible when we interpret “degrees” in terms of probabilistic volatility. We might be uncertain as to what our persistence conditions are, and so we might not know if we could survive the death of our organic brain. Any of us with a healthy open-mindedness about the potential for an afterlife are likely worried about whether our persistence conditions are flexible enough to allow us to come back. But we need a different interpretation of the relevant volatility to make the Persistence Argument logically valid—we have to interpret it as a prediction about how much you are the person in the future event, where this “amount of you” can be increased or decreased at a given time. And it is far less clear that such considerations are any part of prudential rationality.

Indeed, there is an argument that amounts of self at a time should not matter for rational egoistic concern, as long as there is some self to be concerned with. Suppose for reductio you endorsed the following principle:

Synchronic Aggregation: A rational agent will apportion egoistic concern to a self (or selves) at a time based on amount of self (or selves) who exist at that time.

The principle seems plausible enough if the only values for “amounts of self” at issue are determinate and binary: 0 or 1. But matters get trickier if we start to assume that selves can come in other denominations. For instance, how do we assign a self-interested value to an outcome at a time if selves come in increasing denominations?

Retirement Saving with Fission: Suppose I am presented with two investment opportunities. On Opportunity 1: I will pay \$300 into an account now, and I will be allowed to withdraw \$1000 three years later. On Opportunity 2: I will pay \$300 into an account now, enter a personal fission machine once, split into two selves, and each of my daughter selves will receive \$550 dollars three years later.

Synchronic Aggregation seems to entail that I ought to prefer Opportunity 2. On this option I have twice as many selves and, if I aggregate their interests, it seems the investment is as good (from a self-interested standpoint) as the promise of being compensated \$1100 later for a \$300 investment now. This assumes (plausibly, in my view) that rational egoistic concern recommends making short term sacrifices when they will be compensated in the long-run (once probabilities have been taken into account).³¹ Perhaps facts about diminishing marginal utility for money would even further

³¹See Brink (2010, 360-366). Whiting (1986) attributes the compensation account of egoistic concern to Butler and Sidgwick.

recommend Opportunity 2. Each of your daughter selves is likely to get more pleasure per dollar on this distribution scheme.

But this reasoning cannot be correct. The problems get even more stark when the “amounts of self” are allowed to grow further. Suppose I am given Opportunity 3: I will pay \$300 into an account now, enter the fission machine seven times, split into 128 daughter selves, each of whom will receive \$3 dollars three years later. Prudential rationality does not compel me to prefer this option. Indeed, the possibility of fission seems most closely analogous to the possibility of a firm splitting is stock. There can be more denominations, but splitting results in no increase in value.

So either prudential rationality does not recommend making sacrifices now for later, self-involved compensation. Or rational agents will prefer the fissioning options in the cases above. Or—most plausibly—Synchronic Aggregation is false. It might be false because always have at most one determinate self (the option I favor). Or it might be that no-matter how much self we have in reality, we must treat it as one for the stake of assigning value in rational tradeoffs. On either approach, the Persistence Argument for personal discounting is unconvincing.

7 Better Planning through Metaphysics?

Cicero thought that the difference between man and beasts consists in our ability to join up the stages of our life. In *On Obligations*, he writes:

The most important difference between human beings and beasts is this: Beasts, insofar as they are moved by sense, conciliate themselves only to what is at hand and present, since they are aware of very little of the past and future. A human being, on the other hand, shares in reasons, through which he traces consequences, sees the causes of things, notices the mutual relations of effects and causes, compares similarities, and combines and connects future with present things, and so he easily sees the course of his whole life, and prepares the things necessary for living that life.³²

It is a poetic reflection on human nature—unlike beasts, we can intellectually join up the past episodes of our lives, and with a confidence in what we are, we can form plans for the future. It is also sound retirement advice. If we want to live up to our rational commitments, we shouldn’t dwell on how different we’ll be in the distant future. Instead, our rational commitments give us reason to work harder intellectually to “join up” the stages of our predictably volatile lives.

³²Translation from Irwin (2007, 312).

But if personal discounting is prudentially irrational (even inhuman), why is it so tempting? Part of the answer surely has to be the way our emotions influence our ability to perceive our reasons. We can suggest a theory of error for why we have biases in our persistence attributions. For sensible evolutionary reasons, our emotions may be wired to make us deeply short-sighted – feeling great anxiety over pains likely to happen soon and great anticipation at the prospect of impending pleasure. Our ancestors lead short lives, made relatively simple plans, and were very bad at calculating probabilities. These emotions provided a useful heuristic for managing probabilistic volatility, while at the same time disposing us to near bias.

Our emotions are also likely wired to make us sympathetic to individuals we perceive as familiar or morally upstanding. We are wired to be suspicious of those who are different or morally undesirable. Our ancestors needed to find mates, to form tightly-knit social groups, and to quickly distinguish risky individuals in their environment. We needed resilient moral self-regard to project these advantageous qualities to others. These emotions formed a useful heuristic for managing social risk, while at the same time wiring us for similarity and positivity biases.³³

Now we are more complex planners, with longer lives and more sophisticated social networks. Our similarity bias manifests itself synchronically as nefarious and irrational forms of racial bias, ethnic bias, and ageism. It manifests itself diachronically through personal discounting, our lack of concern for our very different older selves. In both cases, these are biases we need to overcome in order to live well and better respond to our reasons.

We live in an age of planning “nudges”. Our banks offer us simple data visualization tools for measuring probabilities and compound interest in our accounts. We implement blind hiring practices to nudge us to disregard arbitrary features about a job candidate’s background. We likewise could use metaphysical nudges to help us avoid personal discounting. For instance, we might favor strong first-personal indexicals when modeling our planning problems (i.e. “how much will *I* need in 2048” rather than “how much will this fund grow” or “what would a future self want”...). And we should be more suspicious both as philosophers and investors of whether our knee-jerk judgments about egoistic reflect stable, coherent assumptions about what we are or what we are worth over time.

³³See Chapter 8 of Sullivan (under contract) for more on this evolutionary theory of error. And see Greene (MS) for more on the relationship between time biases and social biases.

References

- Brink, David O. 2010. "Prospects for Temporal Neutrality." In Craig Callender (ed.), *The Oxford Handbook of Time*. Clarendon Press.
- Cahalan, Susannah. 2013. *Brain on Fire: My Month of Madness*. New York: Simon and Schuster.
- Duckworth, Angela. 2016. *Grit: The Power of Passion and Perseverance*. New York: Scribner.
- Gabler, Neal. 2016. "The Secret Shame of the Middle Class." *The Atlantic* 52–63.
- Gladwell, Malcolm. 2008. *Outliers*. New York: Little, Brown and Company.
- Greene, Preston. MS. "Time Bias, Social Bias, and Value." .
- Harmon, Amy. 2015. "A Dying Woman's Hope in Cryonics and a Future." *New York Times Magazine* .
- Irwin, Terence. 2007. *The Development of Ethics: A Historical and Critical Study*, volume I. Oxford: Oxford University Press.
- Kim, Anne. 2015. "Why You Can't Afford to Retire." *The Washington Monthly* .
- Korsgaard, Christine M. 1989. "Personal Identity and the Unity of Agency: A Kantian Response to Parfit." *Philosophy and Public Affairs* 18.
- Krulwich, Robert. 2014. "Glenn Gould in Rapture."
- Lewis, David. 1976. "Survival and Identity." In Amelie Oksenberg Rorty (ed.), *The Identities of Persons*, 17–40. University of California Press.
- . 1979. "Attitudes *De Dicto* and *De Se*." *Philosophical Review* 88:513–43.
- McMahan, Jeff. 2002. *The Ethics of Killing: Problems at the Margins of Life*. Oxford: Oxford University Press.
- Mischel, Walter. 2014. *The Marshmallow Test*. New York: Little, Brown and Company.
- Nagel, Thomas. 1970. "Death." *Mind* 4:73–80.
- Nakamura, Jeanne and Csikszentmihaly, Mihaly. 2002. "The Concept of Flow." In C.R. Snyder and Shane J. Lopez (eds.), *The Handbook of Positive Psychology*, 89–105. Oxford: Oxford University Press.
- Newman, George E., Bloom, Paul, and Knobe, Joshua. 2014. "Value Judgments and the True Self." *Personality and Social Psychology Bulletin* 40:203–216.
- Parfit, Derek. 1971. "Personal Identity." *The Philosophical Review* 80:3–27.
- . 1984. *Reasons and Persons*. Oxford: Clarendon Press.
- . 1999. "Experiences, Subjects and Conceptual Schemes." *Philosophical Topics* 26:217–270.
- Paul, L.A. 2014. *Transformative Experience*. Oxford: Oxford University Press.

- Sider, Theodore. 2001. *Four-Dimensionalism: An Ontology of Persistence and Time*. Oxford University Press.
- Sullivan, Meghan. 2012. "The Minimal A-Theory." *Philosophical Studies* 158:149–174.
- . under contract. *Time Biases: A Theory of Rational Planning and Personal Persistence*. Oxford: Oxford University Press.
- Tobia, Kevin P. 2015. "Personal Identity and the Phineas Gage effect." *Analysis* 75:396–405.
- Velleman, J. David. 1996. "Self to Self." *The Philosophical Review* 105:39–76.
- Whiting, Jennifer. 1986. "Friends and Future Selves." *The Philosophical Review* 95:547–580.
- Williams, Bernard. 1970. "The Self and the Future." *The Philosophical Review* 79:161–180.
- Zimmerman, Dean W. 2005. "The A-Theory of Time, The B-Theory of Time, and 'Taking Tense Seriously'." *Dialectica* 59:401–57.