

Do We Know How Happy We Are?¹

On some limits of affective introspection and recall

Dan Haybron
Department of Philosophy
Saint Louis University
July 15, 2005

Abstract: This paper aims to show that widespread, serious errors in the self-assessment of affect are a genuine possibility—one worth taking very seriously. For we are subject to a variety of errors concerning the character of our present and past affective states, or “affective ignorance.” For example, some affects, particularly moods, can greatly affect the quality of our experience even when we are wholly unaware of them. I note several implications of these arguments. First, we may be less competent pursuers of happiness than is commonly believed, raising difficult questions for political thought. Second, some of the errors discussed ramify for our understanding of consciousness, including Ned Block’s controversial distinction between access consciousness and phenomenal consciousness. Third, empirical results based on self-reports about affect may be systematically misleading in certain ways.

. . . we are giving you total, dictatorial authority over the account of how it seems to you, about *what it is like to be you*.

Daniel Dennett (1991, p. 96)

By definition, the final judge of someone’s subjective well-being is whomever lives inside that person’s skin. “If you feel happy,” noted Jonathan Freedman. . . “you are happy—that’s all we mean by the term.”

David G. Myers (2000, p. 57)

. . . with respect to his own feelings and circumstances, the most ordinary man or woman has means of knowledge immeasurably surpassing those that can be possessed by anyone else.

Mill, On Liberty

1. Introduction

Never mind what it’s like to be one of Professor Nagel’s bats. I want to know what it’s like to be *me*.² This is a real question, not a joke, because there are good reasons for doubting that any of us have a firm grasp on the quality of our experience of life, in particular its affective character. Possibly, many of us are profoundly ignorant about such matters, to the point that we often don’t know whether we are happy or unhappy, or even whether our experience is pleasant or unpleasant. This, at any rate, is what I will argue in what follows.

¹ Earlier versions of this paper were presented at the 2004 Pacific Division Meeting of the American Philosophical Association, the Consciousness Studies Discussion Forum at the University of Arizona Center for Consciousness Studies, and at Washington University. I am grateful to the audiences, and to Raja Halwani, Tony Jack, Kent Johnson, Joseph Neisser, Eric Schwitzgebel, and Jim Stone for their helpful comments, as well as to David Chalmers for encouraging me to pursue this work.

² The reference is to Thomas Nagel’s famed discussion of whether we can know what it’s like to be a bat (1974).

Consider an illustration. Glen has, for the last twenty years, run a successful machine shop in Detroit. His formative years were spent on his family's ranch in Wyoming, where he thrived. For most of the last few decades, however, he has been fairly unhappy, afflicted by a low-grade but steady mix of depression, anxiety, and stress. Not that he paid it much mind: when his sister, a therapist, first inquired about his feelings, his response was dismissive. "What kind of question is that? Who gives a goddamn how I feel? But if you have to know, I suppose I feel fine. Got nothing to complain about. Yeah, sure, I'm happy." He has since mellowed a bit, growing more receptive to his sister's inquiries, but his basic response never changed much: for the most part he would report, sincerely, that he was happy—although he still didn't think that any sensible person would give much thought to such frivolous matters.

Now Glen finds himself on a ranch near his childhood home, this time for an extended visit with an old friend. The experience is revelatory: back on his old stomping grounds, re-immersed in the activities that engaged him as a youth, it is as if he has transformed into a new person. He experiences real joy for the first time in years, but more than that he feels a tremendous surge of vitality and expansion of spirit. He feels *free* and *big*; by contrast, his usual self, and most of those back in the city, now strike him as tiny, compressed and shriveled up, like ants. He instinctively resumes the confident posture and stride of his youth, and at day's end slips easily into a deep, untroubled sleep. He now realizes that what had previously seemed like happiness was anything but—not because he didn't understand what happiness is, but simply because he was oblivious to the character of his emotional condition. This becomes more apparent still when he returns home to Detroit and gradually resumes that unhappy state. Sadly, the memories of his experience on the ranch fade, and with them the recognition of what he is missing. Once again, he considers himself happy.

I want to suggest that there is a little Glen in all of us. Our powers to assess our own happiness—specifically, our affective states, including their hedonic character—are weaker and less reliable than we tend to suppose. We are, in short, vulnerable to what I will call—for want of a better name—"affective ignorance," or AI. AI can involve two sorts of epistemic failure: ignorance about our past affect ("past-AI"), and ignorance about affects we are currently experiencing ("present-AI"). The central claim of this paper is that widespread, serious errors in the self-assessment of affect are a genuine possibility—one worth taking very seriously. That many of us may be badly mistaken even about our *experience* of life is, I think, an interesting claim in its own right. If we are sound judges of anything, it seems, it would be about what our lives are *like* to us. But we will see that the reflections to follow have a number of other implications for philosophy and for the study of human psychology. For example, they identify a number of difficulties for the use of self-reports in research on well-being and on consciousness (though one aim of this paper is to lend *support* to the use of self-reports by reducing the uncertainties about their limitations).

My arguments also have implications for the understanding of consciousness, including the debate over Ned Block's distinctions among kinds of consciousness. But most importantly, they ramify for our understanding of human welfare and its pursuit: we should be concerned about our competence at assessing, and hence pursuing, happiness. This in turn raises hard questions about the role of the state in promoting well-being.

Few if any contemporary researchers believe us to be perfect judges of our affective states (but notice the optimism of one oft-cited researcher in the epigraph that opens this paper). Yet it is unclear that anyone has recognized just how far off we can be.³ Some of the more extreme forms

³ Two excellent discussions that come close, however, are Schwitzgebel 2002 and Lambie and Marcel 2002. For other helpful discussions of matters closely related to this paper, see Berridge 1999, Berridge and Winkielman 2003, Winkielman and Berridge 2004, and Schooler, Ariely et al. 2003. See also Bargh and Chartrand 1999, Kahneman

of present-AI, in which we are essentially blind to hedonically important aspects of our experience, have yet to gain general recognition, if anyone has noticed them at all. (I find that most people respond with stark incredulity when first told of them.) One of the most prominent researchers in this literature, Timothy Wilson, recently published a book called *Strangers to Ourselves*, detailing how shockingly little we seem to know about our own minds (2002). And yet the chapter discussing AI does not mention that the more serious sorts of errors detailed below are even possible. Moreover, a recent paper by Wilson and Daniel Gilbert defends the accuracy of self-reports of present affective states by claiming that “if [the respondent] is candid and articulate, then one can make the case that his verbal report is unimpeachable.”⁴ Consider also a growing body of work on “unconscious emotion,” including Wilson’s book, that documents affective processes that putatively occur outside of consciousness; the idea that some of these phenomena actually *are* conscious, yet fall outside introspective awareness, seems not, for the most part, to have been taken seriously.⁵ Finally, a couple of the studies cited below as possible examples of present-AI seem to read as though the authors felt it necessary to come up with any explanation of the data *other* than that subjects just didn’t know how they felt. (In at least one case, one gets the distinct impression that the authors suspected precisely this, but feared even mentioning the possibility.) I do not think, then, that what follows will merely be stating the obvious.

While I will cite a variety of empirical results in support of my contentions, these results will often be open to varying interpretations. For the most part, my case will be intuitive. Readers worried about the reliance on intuitive considerations should bear in mind that such arguments form a perfectly respectable part of scientific practice. A standard way of evaluating proposed explanations of experimental results is to ask whether any alternative explanations of the data seem plausible. The list of potential explanations that get taken seriously never depends solely on the data; to some extent we must rely on our best judgment. Accordingly, my goal here is to sharpen our judgment about the possibilities worth taking seriously in this realm.

For convenience, I will assume a more or less hedonistic conception of happiness: happiness consists in pleasant affect.⁶ While I have argued elsewhere against hedonism in favor of a different, “emotional state” view of happiness, we can set aside the differences here.⁷ Readers who

1999, Nisbett and Wilson 1978, Simons and Chabris 1999, and Zajonc 1980, 1994. For those unfamiliar with the “basketball” video used by Simons and Chabris, I highly recommend viewing it *before* someone gives away the surprise. It can be downloaded, with instructions, at <http://viscog.beckman.uiuc.edu/media/ig.html>. I’m not sure if this version is large enough to obtain the effect, but suffice it to say that most viewers find the result astonishing.

⁴ Gilbert, Pinel et al. 2002, p. 308. The line originally appeared in Gilbert et al. 1998. The following paragraph contains the line, in reference to their subjects’ self-reports of happiness, “whether the Experiencers’ beliefs were right or wrong is a question to which no one—Experiencer, philosopher, or psychologist—can at present offer a definitive answer.” It is not entirely clear how to read this, since self-reported happiness is not generally thought to be simply a report of occurrent affect (and would be only marginally interesting if so, since one might feel happy at a given moment without *being* happy at all). I suspect it strikes a more cautious note precisely because these reports go beyond immediate feelings.

⁵ See, e.g., Berridge 1999, Berridge and Winkielman 2003, Winkielman and Berridge 2004, and Prinz 2004. A noteworthy exception is Lambie and Marcel 2002.

⁶ My discussion will apply only indirectly to life satisfaction views of happiness, which typically focus on agents’ global evaluations of their lives (see, e.g., Nozick 1989, Sumner 1996). AI is nonetheless significant for determining how seriously we should take attitudes of life satisfaction. For if we don’t know how pleasant our experience is, we can hardly expect to make informed judgments about the satisfactoriness of our lives. I discuss life satisfaction further in Haybron forthcoming-a, forthcoming-b.

⁷ While not crucial for our purposes, these differences are not at all trivial. In fact AI is substantially *worse* on the emotional state view. My sole nod to the emotional state theory is to focus on “affect” more broadly, rather than simply talking of pleasant and unpleasant experiences. See Haybron 2000, 2001b, 2003, forthcoming-b.

object to this usage of ‘happiness’ can call the psychological states I discuss anything they like; little of substance hangs on the word.

I will not try to catalogue every possible source of AI, but will focus only on some of the more interesting varieties. Thus, for instance, I will set aside worries about self-deception, as well as problems of misidentification, where subjects are fully aware of their affects and their phenomenal qualities but classify them incorrectly (say by misapplying the concept *upset* to a case of anger).⁸ Nor will I discuss problems with self-reports not bearing on AI, for instance when subjects edit their reports to suit the social circumstances. We begin with present-AI.

2. Ignorance about present affect

2.1 Elusive affects

Paradigm affects like being gripped by terror, writhing with pain, or giddy with elation seem pretty hard to miss. We probably are pretty good judges, at least in broad terms, about these and many other affects. And we seem to be particularly observant immediately following a sharp change in affective state, where the contrast highlights qualities of our states that may otherwise be less obvious. (This phenomenon plays a significant role in a number of the arguments to follow.) But some affective states are more elusive than the paradigmatic ones, particularly moods and mood-like states such as anxiety, tension, ennui, malaise, or the pleasant state of “flow” that comes with losing oneself in challenging and rewarding activities.⁹ They may exceed our powers of discernment even while they are occurring.

How is this possible? Consider, first, why paradigm affects like strong emotions or intense pains seem to be so hard to miss. Most obviously, they are *intense*. But more interestingly, they are relatively focused states, having a more or less specific object or (phenomenological) location: one is terrified of the spider crawling up one’s leg, elated over getting the job of one’s dreams, or feeling pain in one’s freshly stubbed toe.¹⁰ Such affects are thus comparatively easy to attend to. To discern a pain in your toe, for instance, you need only attend to your toe—and even this will typically happen automatically, given that stubbed toes have a way of calling attention to themselves. Similarly, if you are feeling uneasy about visiting the dentist, you can attend to your unease simply by thinking about your impending appointment and noting how that feels—something that, again, tends to happen all by itself. But if you are feeling uneasy *in general*, about nothing in particular, to what do you attend? This feeling attaches to nothing in particular; it just forms a part of the background to your experience of everything. You must, then, attend to a diffuse background phenomenon that pervades every aspect of your experience—not impossible, but a bit like attending to the air around you. There just isn’t much to grab hold of.

The same seems true of moods in general (and perhaps other affects as well, but for convenience I will focus on moods). Unlike emotions and sensory affects (*viz.*, physical pleasures and pains), moods have no particular location or object. (If they have an object at all, it is everything,

⁸ Eric Schwitzgebel has recently made some interesting observations along these lines (2002); see also the discussion of alexithymia in Lambie and Marcel 2002, p. 251. By misidentification I have in mind issues akin to linguistic failures, though the differences between this and some of the problems of AI that I discuss below can be difficult to make out. Thanks to Joseph Neisser for bringing this worry to my attention.

⁹ On the notion of flow, see Csikszentmihalyi 1990.

¹⁰ “Focused” is vague, and it is tempting to put the point in terms of intentionality. But the intentionality of sensory affects like pains is controversial. And those who think pains purely sensory states without intentional content can still grant that our perception of pain is typically of something with a fairly definite location. Hence “focused.”

which phenomenologically is pretty much like having *no* object.¹¹) They are also highly diffuse, pervading the whole of one's consciousness. They are, moreover, comparatively diaphanous, offering us not so much distinct objects within the field of consciousness as alterations of the field itself, coloring the entirety of our experience. Finally, moods tend not to be very intense. They are usually low-key affects that do not call attention to themselves the way that, say, sensory pains do.

Yet moods are quite central to the experienced quality of our lives, far out of proportion to their grip on our attention. A vague sense of malaise might easily go unnoticed, yet it can sour one's experience far more than the sharper and more pronounced ache that persists after having stubbed one's toe. Likewise for depression, anxiety and related mood states, at least in their milder forms. Consider how a tense person will often learn of it only when receiving a massage, whereas stressed or anxious individuals may discover their emotional state only by attending to the physical symptoms of their distress. Presumably being tense, anxious, or stressed detracts substantially from the quality of one's experience, even when one is unaware of these states.

So elusive are some moods and mood-like states that whole classes of them may fail to register in our collective awareness. Many distinctions in moods surely remain to be made, but two in particular are worth noting. The term "flow," for one, has yet to enter the popular vocabulary, but we can all recognize this important phenomenon once it is described. This, as I noted earlier, is a state of total absorption in an activity, normally an activity that is challenging yet not too difficult, and which is perceived as worth doing. (And which is being done well.) The athlete experiences flow when she is "in the zone," as does the novelist lost in his work. Flow is often described as having no phenomenology, since we are not aware of feeling anything while it happens. (Indeed, thinking about how one is feeling typically ruins the experience. See Schooler, Ariely et al. 2003.) Yet it clearly does feel like something, since it is manifestly a pleasant state. Indeed, some researchers argue that it is a crucial element of happiness.¹² But the very existence of flow has largely gone unrecognized. Talk of things like being in the zone suggests a faint awareness of the phenomenon, but the very absence of a word indicates just how faint that awareness is. (Aristotle's discussion of pleasure in the *Nicomachean Ethics* has clear resonances with the notion of flow. But again, the concept itself seems not to arise in his work.)

A related phenomenon arises in the case of Glen. Recall the "expansion of spirit" he experienced when visiting his friend's ranch: a sense of freedom and inner enlargement, of being uncaged and allowed to grow to full size, of being *alive*. His workaday existence in Detroit, by contrast, left him feeling "pinched and hidebound,"¹³ tiny and antlike, drawn-in and shriveled up like a pea, dead—compressed. Call this *affective compression*, and its more positive counterpart "expansion." I am not sure how to characterize this state more precisely, though it appears to be strictly a property of a person's emotional state rather than a distinct type of affective state. But I hope that the reader will recognize it from the crude sketch I have provided. The contrast might be seen in comparing our stereotype of, say, an English aristocrat with that of a flamboyant Caribbean islander.¹⁴ Compression has obvious affinities to depression, though anxiety and other affective states seem to involve it as well. (Perhaps compression is actually a mild form of depression.)

¹¹ Recent philosophical accounts of mood include Armon-Jones 1991, Griffiths 1997, Lormand 1985, 1996, Prinz 2004, and Sizer 2000. I differ from some of these authors in holding that moods typically have non-dispositional, phenomenal aspects. Surely there is something it is like to be in most moods, and quite apart from whatever other affects the mood generates. For overviews of scientific work on moods, see Ekman and Davidson 1994, Frijda 1993, Morris 1999, and Parkinson et al. 1996.

¹² See Csikszentmihalyi (1990, 1999).

¹³ To borrow an expression from Mill's essay on individuality (1991).

¹⁴ These stereotypes may, of course, be unrepresentative. While we're trading in caricatures, an entertaining example of the uncompressed person can be seen in Jack Palance's flinty cowboy, Curly, from the movie *City Slickers*.

In any event it isn't pleasant, and while it rarely gets noticed it is arguably quite common. Indeed, those who have undergone experiences like Glen's might see compression as more or less ubiquitous in our society.

Why would moods and related affective states be so elusive? This is no place to develop a theory of moods, but a plausible conjecture is that moods function to govern our responses to the general circumstances of our lives, whereas emotions and sensory affects constitute responses to specific stimuli.¹⁵ Suppose this is correct. Thus, for instance, anxiety may serve in hostile circumstances to make us vigilant and primed to respond to threats. For this purpose it is not essential that anxiety command our attention. On the contrary: its very point is to keep our attention available to focus on whatever problems may arise, wherever they occur. The disgust felt when drinking spoiled milk, by contrast, will prove fruitless if it does not focus the mind rather sharply on the issue at hand. Because it is so narrowly directed and involves one's full attention, the feeling of disgust should not be particularly hard to notice. A generalized anxiety, being utterly diffuse and leaving one's attention free for other matters, may be comparatively easy to overlook—precisely because it is supposed to be.

Let me take this a step further and suggest why we might *expect* the chronically anxious to seem like, and think, they're happy. The reason is that anxiety may be able to do its job of making us more vigilant and so forth without rational processes being aware of it at all—indeed, with the agent thinking everything's just fine. If so, then it may be adaptive in many situations to be anxious while *believing* that one is happy, and more generally *presenting* oneself as happy. For overt displays of anxiety project weakness and vulnerability, whereas displays of happiness send the opposite message. It isn't hard to see the advantage, in many circumstances, of being highly prepared to deal with threats while projecting a relaxed and confident demeanor. (Perhaps Americans are experts at this.) Similar remarks may apply to a variety of other elusive states, like being stressed, or depression or compression, and perhaps mood in general. In short, some problems are best handled by subrational processes, leaving reason out of the loop if not downright misled.¹⁶ If this is right, then such states may not serve the informational function performed by emotions and sensory affects. (Note that this would eliminate an important rationale some have posited for hedonic adaptation in relation to such states: that affects must diminish with time to free up attentional and other cognitive resources. See, e.g., Wilson and Gilbert 2003.)

An interesting further conjecture is that elusiveness should tend *not* to attach to affects associated with universal facial expressions like the “basic emotions” of sadness, feeling happy, anger, fear, contempt, disgust, and surprise.¹⁷ Such emotions clearly serve important signaling functions: they try to make themselves known, at least to others. Plausibly, it would be most adaptive for such emotions to be known to their bearers as well. If you are displaying anger to a big fellow

¹⁵ See, e.g., Lazarus 1994, Morris 1999, Prinz 2004. I argue for this idea in Haybron 2001a.

¹⁶ See Bargh and Chartrand 1999. Someone might object that if certain affects don't serve an informational function, indeed aren't supposed to come to our attention, then why should they have phenomenal qualities at all? Good question; here are two possible responses: one, moods do their work partly by altering how things seem to us—making things seem more threatening, say—and thus affect our experience as they do; second, elusive affects might serve an *attenuated* informational function, being designed mostly to work unnoticed, or at the periphery of awareness, so as to free up cognitive resources, but prone enough to grab our attention that we won't normally be forever in the dark about what's going on.

¹⁷ See, e.g., Ekman 1994. However, there is some evidence that even these emotions can be elusive. E.g., one study had sufferers of a spider phobia engage in a series of increasingly difficult interactions with spiders, from walking towards a spider in a jar to letting it walk on their hands (Arntz 1993). Those who received naltrexone, an opioid antagonist, refused to proceed with the tasks at a significantly earlier stage than those who received a placebo, indicating an enhanced fear response. Yet self-reports of fear in both groups were essentially the same at all stages. It is not implausible that naltrexone recipients experienced heightened fear but failed to notice it.

before you, for instance, you might want to be clued in. Perhaps not coincidentally, then, the elusive states we have discussed do not seem to involve canonical facial expressions. They seem not to have the kind of signalling function served by the basic emotions. Maybe, at least for some of them, their work is best done in secret.

These remarks are of course highly speculative. But they are not implausible, and seem broadly consistent with what we know about mood.

2.2 Adaptation to persistent affect

Everyone knows that we often adapt to things over time: what was once pleasing now leaves no impression or seems tiresome, and what used to be highly irritating is now just another feature of the landscape. Could it also be that some things are lastingly pleasant or unpleasant, while our *awareness* of them fades? I would suggest that it can.

Perhaps you have lived with a refrigerator that often shrieked due to a bad bearing. If so, you might have noticed that, with time, you almost entirely ceased to notice the racket.¹⁸ But occasionally, when the compressor stopped, you *did* notice the sudden, glorious silence. You might also have noted, first, a painful headache, and second, that you'd had no idea how obnoxious the noise was—or that it was occurring at all—until it ceased. But obnoxious it was, and all the while it had been, unbeknownst to you, fouling your experience as you went about your business. In short, you'd been having an unpleasant experience without knowing it. Moreover, you might well have remained unaware of the noise even when reflecting on whether you were enjoying yourself: the problem here is *ignorance*—call it reflective blindness—and not, as some have suggested, the familiar sort of inattentiveness we find when only peripherally aware of something.¹⁹ In such cases we can bring our attention to the experience easily and at will. Here the failure of attention is much deeper: we are so lacking in awareness that we *can't* attend to the experience, at least not without prompting (as occurs when the noise suddenly changes).

There is some empirical evidence that persistent noise can degrade the quality of our experience in ways that escape our notice. In a recent study of office noise, physiological and behavioral measures of stress were significantly higher in subjects who had worked for three hours in a simulated office with low-intensity noise (55 dBA) than in subjects assigned to a quiet office (40 dBA). Noisy-office subjects showed elevated epinephrine levels, made half as many ergonomic adjustments to their workstations, and were markedly less persistent in efforts to solve difficult puzzles afterwards. Yet the researchers were surprised to find *no* differences in reports of perceived stress—specifically, reports about the extent to which subjects felt “bothered, worried, relaxed, frustrated, unhappy, contented, [or] tense.”²⁰ Plausibly, the noisy-office subjects had adapted somehow to the conditions, and saw nothing amiss. By contrast, an earlier study involving a much briefer noise exposure, and subjects who were much less experienced at office work, *did* find differences in perceived stress. It seems likely that given enough time, the experienced office

¹⁸ Interestingly, Ned Block has been using essentially the same example in his work Block 1995; his use is independent and predates mine. Evidently the experience is not idiosyncratic.

¹⁹ See, e.g., Rosenthal 2002b, Chalmers 1997, Church 1995, Prinz 2004, and Searle 1990. The fact that we can come to know about an affect under certain conditions, like after a sudden change, does not affect the point—*viz.*, that *as things are*, we are unable to discern it.

²⁰ Evans and Johnson 2000. Consider also the experience of a dull persistent pain; for related findings with anxiety, see Lambie and Marcel 2002. Frederick and Loewenstein report that noise is actually one of the things to which people do *not* seem to adapt (1999). This is surely correct for some types of noise. But it is equally clear that adaptation, at least to steady noise, can occur (see, e.g., Glass and Singer 1972, Glass, Singer et al. 1977). (There are also multiple forms of adaptation—physiological, hedonic, etc.) Frederick and Loewenstein also express skepticism about adapting to depression and pain *qua* unpleasant experience, but they do not address the questions raised here.

workers used in the newer study ceased to notice the unpleasant effects of the noise. Yet it also seems plausible that the noise affected not only subjects' physiological responses and behavior, but the hedonic quality of their experience as well: they experienced more stress, had a less pleasant time of it, than they would have without the noise.²¹ Other interpretations of the data are of course possible. But setting aside matters of health, *which group would you prefer to be in?* Would you really be indifferent? (Cases like this make a good test of the skeptic's convictions, since matters of welfare are at stake. Those who deny that we can have phenomenality without knowing it should presumably be indifferent to such a choice.)

Attentional adaptation, as we might call it, probably extends to persistent affect quite generally, at least where the affect is not especially intense. Take two people, one who is newly depressed after a long period of happiness, and another who is equally depressed, but has been like this for some years. Is it really plausible that both individuals will judge their experience equally unpleasant? Rather, the freshly depressed person will probably be much more vividly aware of the awfulness of her predicament; the long-depressed individual, by contrast, will likely have gotten more or less used to his suffering, and perhaps no longer notice it at all. Perhaps he has even come to regard his depressive baseline as a normal, not unpleasant, state of being.²²

Much has been written in recent decades about a "hedonic treadmill": we tend to adapt to changes in our lives, so that we tend eventually to wind up no more or less happy than we were to begin with.²³ Such phenomena are so striking in many empirical studies that some investigators regard us as having happiness "set points": characteristic levels of happiness to which we tend to return almost regardless of what happens in our lives. These set points are said to be substantially heritable, perhaps as much as .80.²⁴ But given the heavy reliance of these findings on self-report measures, and given the likelihood that adaptation is sometimes at least partly attentional rather than hedonic, these findings may prove to be exaggerated. In short, we may be substantially on an attentional treadmill instead of merely a hedonic one.²⁵

2.3 Attentiveness and discernment

There is no reason to suppose that we are all perfectly, or equally, attentive and discerning observers of our own affect. Surely the Dalai Lama notices things about his emotional state that would have completely escaped John Wayne. We can't all be introspective Prousts, and probably very few of us are. The problem that arises here is not simply that some affects are elusive, or become so through adaptation, but that a given affect can elude some people and not others. And in-

²¹ It is possible that subjects' experience was degraded in ways that did not escape their notice, but simply weren't captured by the self-report instrument. For instance, we need to distinguish between the unpleasantness of the *noise* and the unpleasantness of the *affect* (e.g., the stress response) it causes. Perhaps the study missed differences by asking only about stress-related unpleasantness. But this seems unlikely: if you perceive a noise to be unpleasant, you will probably find it unpleasant in ways that are covered by the seven-item instrument in question. Finally, given the subtlety of the feelings involved, it is also possible that differences were obscured by the use of a coarse-grained, four-point scale to elicit self-reports. Perhaps a 100-point scale would have yielded different results.

²² We must make allowances for the extra distress that the first individual might feel at the recognition of her plight. But even setting that aside, it is implausible that the long-depressed individual will be as fully aware of the unpleasantness of his state as the newly-depressed person.

²³ See Headey and Wearing 1989, 1992, Diener and Diener 1996, and Suh, Diener and Fujita 1996.

²⁴ See, e.g., Lykken and Tellegen 1996 and Lykken 1999.

²⁵ Kahneman has made a related suggestion regarding a "satisfaction treadmill" (1999, 2000b). However, the mechanism he suggests for this involves not fading awareness but escalating aspirations: we need more pleasure to be satisfied or to consider ourselves happy. In an excellent paper inspired largely by the literature on adaptation, Elijah Millgram has argued that happiness serves the modest function of indicating *changes* in well-being (2000). His case weakens, however, insofar as claims about adaptation have been overstated.

sofar as one is relatively inattentive about one's affective state, or an undiscerning observer of what one does attend to, one is liable to be comparatively ignorant about the quality of one's affect.²⁶

This sort of problem may be particularly acute for those raised in cultures that discourage emotional introspection. I have been told that my relations of a few generations back would not have *understood* an inquiry into their emotional states, much less figured them prominently in their assessments of their lives. I do not know whether this is true, but the preoccupation of present-day Americans with their feelings is clearly more pronounced than it was in the pre-war years, perhaps to a fault. It would not be at all surprising to find that self-reports of affect among Americans are more accurate now than they used to be. They may also be more accurate than those of individuals from so-called "collectivist" cultures, such as those of many Asian countries. Such cultures place far less emphasis on emotional states inasmuch as communal concerns are believed more significant than personal ones.²⁷ Thus judgments of life satisfaction in collectivist cultures have been found to draw less on affective state than judgments in individualistic cultures, and in some countries—like China—may not correlate with negative affect *at all* (Suh, Diener et al. 1998). It is hard to believe that people in any culture are *that* indifferent to how they feel. Perhaps people in countries like China are simply less astute observers of their inner lives than Americans are, precisely because they assign a lower priority to such matters.

Consider a study that compared self-reports of emotion with autonomic nervous system (ANS) activity patterns among American and Indonesian (specifically, Minangkabau) subjects after they had been instructed to engage in various emotional expressions such as smiling and frowning (Levenson, Ekman et al. 1992). The American subjects were much more likely to report experiencing the corresponding emotions, even though ANS profiles were the same between the two groups. It is possible that the Minangkabau refused to apply the emotion concepts in the absence of the usual eliciting circumstances, a hypothesis offered by the authors. Yet the fact that some Indonesians *did* report emotion experiences indicates that their emotion concepts permit emotion ascriptions in the absence of appropriate eliciting factors. A more plausible interpretation, I would suggest, is that the Minangkabau *had* the relevant emotional experiences, but weren't able to identify them without the usual external cues.

2.4 Scale norming

How pleasant would you say your experience is, on a scale of one to ten? Your answer to this question will depend, not only on what you take your experience to be, but also on where you draw the lines between good, neutral, and bad, as well as on the range of possible experiences you are comparing it against. What counts as a ten, or a one? Different people will answer differently, and your answer today may differ from the answer you would have given a few years ago. A Manhattan debutante may consider a day without her cell phone about as bad as it gets, while a crippled resident of Manila's trash dumps is liable to have rather different standards of unpleasantness. In short, people can use widely varying scales when rating their experience, a phenomenon known as *scale norming*. Thus one person's ten might be another's six.²⁸

Scale norming may strike the reader as a mere technical issue regarding measurement techniques. But its existence has nothing to do with empirical research methods, and generates a quite serious—and probably widespread—form of AI, namely ignorance about the *relative* quality of our experience. For one thing, it undermines interpersonal—and *intrapersonal*—comparisons inso-

²⁶ See the discussion of alexithymics in Lambie and Marcel 2002, and Schwitzgebel on introspective training 2004.

²⁷ For a good collection of scientific papers on cultural issues and happiness, see Diener and Suh 2000.

²⁸ See Frederick and Loewenstein 1999. The phenomenon is not of course limited to numerical scales.

far as we aren't sure whether two individuals, or two time-slices of the same individual, are rating their experience using similar scales. More importantly, scale norming suggests that we often fail to recognize how our experience rates in relation to the possibilities. To the extent that this happens, we don't know what we're missing, and therefore don't know how good (or bad) what we've got is.

Consider someone who has been more or less depressed for years. Even if she knows she's depressed—a large “if”—she may well be so deprived of great pleasures that she simply cannot comprehend the vastness of the gap between her present state and happiness. She may, in other words, fail utterly to recognize how unpleasant her lot is. Conversely, someone who has led such a sheltered life that he has had few occasions for real suffering may have little comprehension of just how pleasant his experience really is. Probably all of us have experienced corrections of such misapprehensions, for instance when visiting a hospital.

As these two examples suggest, scale norming can be a source of adaptation. Insofar as this happens, then we have a second reason for thinking that hedonic adaptation may be less extensive than self-reports would suggest: for these reports may reflect a kind of “scaling adaptation,” in which subjects' scales change over time. Lasting changes in hedonic state may thus be obscured by adjustments in the scales we use to rate our experience.

2.5 Affect-type bias

Affect-type bias has received little if any attention, but may profoundly impact the quality of our self-assessments. The idea is that some types of affect are more likely to be considered than others when judging the quality of our experience. When judging how happy we are, for instance, we may focus primarily on affects that fall along the joy-sadness dimension rather than on those, say, along the anxious-calm dimension. And we may do this even if we believe that the latter sorts of affect are just as important for happiness than the former. Such biases might reflect societal or personal tendencies to regard certain affects as more representative of the affective realm, of happiness, or simply as more important (but for reasons other than their hedonic quality). Alternatively, biases might arise simply because some affects are more elusive than others. Classic emotions like joy, sadness, fear, or anger tend to be easily identified and labeled, and hence may be relatively likely to be considered in our reflective assessments.²⁹ We probably cannot say the same about more diffuse states like anxiety, tension, ennui, malaise, flow, or compression. One possibility worth taking seriously is that such biases frequently cause self-reports to overlook states, like anxiety and stress, that are crucial to the experienced quality of our lives—even if we are aware of them while they are happening, and even if we are reporting on our immediate experience.³⁰ Inasmuch as stress is often said to be one of the hallmarks of our age, our primary measures of happiness may be, at least in that respect, overly optimistic.

Affect-type biases, along with problems of elusiveness and scale norming, may constitute a third AI-based source for some of the findings on adaptation. It is possible, indeed probable, that hedonic adaptation is more pronounced with some types of affect than others. Perhaps, e.g., we

²⁹ This differs from simple elusiveness in that it concerns not failing to be aware of certain affects but failing to take them into account when making judgments.

³⁰ One study, for example, found that anxiety and tension correlated with life satisfaction at only -0.17, whereas feelings of depression and dejection correlated much more strongly, at -0.46 (Pilcher 1998). This might simply reflect differences in how much subjects cared about these affects, though the differences are rather large. In fact the study instruments *themselves* sometimes incorporate affect-type biases. E.g., the popular Fordyce emotions questionnaire only assesses joy-sadness-type affects (Fordyce 1988), and so cannot be expected to measure other dimensions of happiness effectively. I argue that joy-sadness affects are far less important than widely believed for being happy in Haybron forthcoming-b.

adapt more with states of joy and sadness than with states of anxiety or tranquility.³¹ If so, and if we are also less likely figure the latter states in self-reports of happiness or emotional state, then we might appear to adapt completely to circumstances that in fact leave us lastingly more anxious. In short, the adaptation research might be skewed by an undue focus on certain types of affect at the expense of others. (None of this is to deny that hedonic adaptation is real and significant. But it does indicate that happiness might not be as immutable as many have supposed.)

2.6 Expectation effects

To the extent that we find it difficult to render judgments about how happy we are, or about the quality of our experience, we might rely somewhat on our *expectations* concerning how we should feel, or are likely to feel, given the circumstances. A groom on his wedding day might fail to come to grips with the complexity of his emotions and think himself filled with pure joy, mainly because that's how he is supposed to feel, and how he imagines any newlywed must feel. Similarly, if a researcher comes and asks you, Bill Gates, how happy you are, you might sincerely claim to be *very* happy, in great part because you are, after all, the wealthiest human being on the planet. So naturally you must be happy—who wouldn't be?—and insofar as your own introspection doesn't yield obvious answers one way or the other, you may conclude that, yes, you are indeed quite a happy fellow. And someone who just lost a pet might overrate her unhappiness because she came into the situation believing she would, or should, be miserable.

Such reasoning need not be explicit, of course; more likely it would involve the sort of processes that occur when we confabulate memories of past affect.³² Indeed, expectations are what *drive* such confabulations, and expectation effects clearly impact our recall of past affect (and hence can be a source of past-AI as well).³³ What I am suggesting is that they can affect judgments about the present as well. Consider also that we often infer how we feel by observing our own behavior and physiology, as when one recognizes an irritable mood as the best explanation of having put one's computer out the window.³⁴ Such guesswork will obviously be subject to our expectations or "theories" about affect.

2.7 Can we tell pleasure from pain?

It is one thing to claim that we can be wrong about *how* pleasant or unpleasant our experience is. But might we even get the *valence* of our current experience wrong? That is, might we mistake pleasant for unpleasant states, and vice-versa, even while they are happening? It seems possible that we could, at least with respect to moods. Problems of elusiveness, scale norming, and adaptation might cause us to become so profoundly mistaken about the basic character of our experience that we can't even tell whether it is pleasant or unpleasant. Such mistakes are clearly at least conceivable, given just the distinction between having an experience and being reflectively

³¹ Recall the earlier suggestion that claims about the need for affects to fade with time to free up cognitive resources may not apply to states of the latter sorts.

³² For references, see Wilson and Gilbert 2003.

³³ The term 'expectation effect' may be a bit misleading given that researchers use it to characterize the way people's expectations of a future event can affect their experience of it. But talk of 'theories' of affect, a more common way of describing these issues, fails to encompass normative beliefs about what *should* occur, and it seems to me that these can affect our judgments similarly. It also seems a bit too intellectualized.

³⁴ Schooler et al. argue that self-appraisals of affect draw heavily on inferences like this (2003). Cases like the one just described may sometimes involve a sort of irritability that doesn't feel like anything at all, since an irritable mood could be little or no more than a disposition to overreact to things (Haybron 2001b). But I am imagining a case where one really does feel irritable, but doesn't recognize it until led to do so by inference from one's behavior.

aware of it, along with the (in principle) fallibility of reflection.³⁵ Less clear is whether human psychology could actually give rise to them.

Yet this is precisely what appears to happen in Glen's case: though unhappy, he sincerely judges himself to be happy. What might explain this? For starters, his judgments—I will suppose—tend to exhibit a positive bias:³⁶ despite his gruff demeanor, he has little patience for whining or complaining and tends to look on the positive side of things more than the negative. As a result, many of the negative affects that he experienced are simply overlooked. Second, his unhappiness consists largely in some of the more elusive affects, such as emotional compression, or being mildly anxious or stressed. He simply never notices them. Adaptation, moreover, has caused him to stop noticing other affects, such as his depressed mood. And affect-type biases lead him to ignore even many of those affects that do fall within his awareness. Nor does it help that he is pretty near the opposite of a Proust as an observer of his inner state: he tends not to think about it at all, and when he does attend to it his perception is crude and undiscerning, like a five-year-old hearing Stravinsky. Finally, the relative homogeneity of his recent experience has caused him to assess it on a truncated scale: what now seems to him like a highly desirable condition would, on the ranch, strike him as a pale imitation of happiness. It seems plausible that, under such conditions, he might believe his experience to be pleasant even when it is unpleasant. And this could be just what he realizes when he suddenly finds himself truly happy. If so, then he has mistaken an unpleasant experience, indeed an unpleasant *life*, for a pleasant one.

People do sometimes take themselves to have had experiences of this nature. (I am one of them.) The only question is whether their retrospective assessments should be taken seriously.³⁷ Given the many ways such evaluations can go wrong, we should certainly be cautious here. But it nonetheless seems plausible that Glen got the very valence of his experience wrong. Similarly, a chemotherapy patient's good day might be a newlywed's bad day—that is, the very experience that one considers pleasant may be judged unpleasant by the other.³⁸

It may seem absurd that we could possibly be unable to tell pleasure from pain. Even if we grant the conceptual possibility, such ignorance might seem utterly alien to human psychology. With respect to pronounced sensory pleasures and pains, as well as emotions, this might be correct. For in such states the pleasure or pain attaches to an object, which makes it easier for us to attend to our feeling. Moreover, we presumably will feel either aversion or attraction (or some other pro- or con-attitude). Thus one way of telling that the fire feels bad is that I very much want to pull my hand out of it. It also helps that the experience is intense and highly localized. But a vague and generalized anxiety, by contrast, can be so low-key, diffuse, and free-floating that it is harder to say how we know its valence. To what do we feel aversion or attraction? Everything? The point is not that moods are altogether hedonically inscrutable. It is just that they offer fewer and less obvious cues as to their hedonic character.

³⁵ Well, this assumes that (un)pleasantness doesn't *entail* reflective awareness. Someone might object that pleasant and unpleasant experiences entail having pro- or con-attitudes such as liking or wanting. For instance, attitudinal theories of pleasure identify it roughly with having an experience that one likes or wants, and thus would seem to have this consequence (see, e.g., Sumner 1996, Feldman 2004). But even if pleasure always involves attitudes of wanting or liking, these attitudes need not always be introspectible. Zajonc and others, e.g., have found that we are constantly evaluating our experience, but that the relevant attitudes occur automatically and unconsciously, at a very primitive psychological level (Zajonc 1980, 1994; see also Bargh and Chartrand 1999).

³⁶ I.e., he is what I will call a "Pollyanna." I discuss this notion in §3.3.

³⁷ Schwitzgebel discusses a case of a woman who consistently claims to find gardening unpleasant, evidently sincerely, yet manifests all the signs of joy while doing it (Schwitzgebel 2002).

³⁸ This is an example given by Frederick and Loewenstein 1999, who suggest that scale norming can affect the reported valence of a person's experience.

3. Ignorance about past affect

3.1 Omission

Errors about the past are less striking than those concerning the present, since no one expects us to have perfect recall. Yet such mistakes can leave us substantially blind about our welfare, since in evaluating our lives we have very little else to go on but our memories, the present being the fleeting thing it is. The most mundane source of past-AI is simple omission: we experience countless affects every day, and there is simply no way to bring all of them to mind when reflecting on the past quality of our experience. Many affects we have completely forgotten about, while others remain in memory but do not come to mind when making a judgment about our experience. Errors of omission should not prove worrisome as long as we do not omit too much, and are evenhanded about the affects we omit. Yet there is some evidence that we do tend to omit quite a lot, and that our retrospective assessments are anything but unbiased.³⁹ Affect-type biases and expectation effects, for instance, may be even more prominent when recalling past affect than present, since the relevant information is less accessible.

3.2 Peak-end effects and duration neglect

A number of studies have found that the remembered quality even of a recent experience is strongly biased by the nature of its peak moment—where the intensity was greatest—and by its ending. Indeed, the remainder of the experience seems largely to be ignored. Such “peak-end” effects, as they are called, can cause people to form preferences for *less* pleasant experiences over more pleasant ones (Kahneman, Fredrickson et al. 1993, Kahneman 1999, 2000b, 2000a). For instance, doctors can improve the remembered quality of a painful medical procedure (a colonoscopy) by extending the procedure, but with a slightly less painful coda. That is, you can make an experience seem less painful in retrospect by *adding more pain*. Not only did this manipulation yield surprising memories of hedonic quality, it also affected future choices: patients given the more painful procedure were more likely to return for another exam five years later.⁴⁰ Similar results have been found with a wide range of situations: the peak-end rule appears to govern hedonic recall quite generally.

A corollary of the peak-end effect is that our memories of hedonically salient experiences tend to disregard the length of the episode—*duration neglect*. For instance, a series of colonoscopy studies varied between 4 and 69 minutes in duration; yet the correlation between the duration of the procedure and subjects’ retrospective evaluations of it was only .03.⁴¹ Schwarz and Strack suggest in light of this phenomenon that three years of economic hardship may not seem much worse in retrospect than one year if the peaks and ends are about the same (1999). Needless to say, this makes retrospective assessments of experience problematical.

³⁹ For references, see Wilson and Gilbert 2003. Concerning blank omission, one striking finding indicates that many individuals diagnosed with depression forget about it not long after the episode passes. In one study, 37 percent of depressed subjects denied having been depressed when interviewed six years later (Warshaw, Klerman et al. 1991). It is possible, however, that they nonetheless remembered having a really bad experience, but did not label it as depression. Interestingly, subjects with the most recent depression exhibited the worst recall. From a hedonic standpoint, this is a bit like forgetting that one had cancer.

⁴⁰ Redelmeier, Katz et al. 2001, cited in Kahneman and Frederick 2002.

⁴¹ Redelmeier and Kahneman 1996, cited in Kahneman 2000a.

3.3 Affect-valence bias

The peak-end effect is the most striking type of bias to have been documented by empirical researchers. But at least three further types of bias are worth distinguishing. The first, *affect-valence bias*, concerns subjects' tendencies to focus disproportionately on either negative or positive affect.⁴² "Kvetches," for instance, are more likely to draw on negative experiences. We have all known individuals who seem pretty happy but invariably launch into a litany of woes when asked how they are doing. "Pollyannas," by contrast, focus on the positives. There is considerable evidence that most people are Pollyannas, exhibiting a wide range of positive biases in memory recall and judgment.⁴³ Lykken and Tellegen, for instance, found that over 86 percent of subjects rated themselves more "happy and contented" than about two-thirds of people (1996). And people tend to recall positive past events and emotions more easily than negative ones.⁴⁴ Insofar as we are Pollyannas, then, we are probably less happy than we think we are (and conversely for Kvetches).

3.4 Present-affect and situational biases

Two further types of bias concern the context in which judgments are made. With *present-affect bias*, the subject's current affective state influences which types of affect are most likely to come to mind. In particular, we tend to recall affects that are congruent with our present state more than those that are not.⁴⁵ When in a good mood, for instance, you are more likely to recall similar moods from the past rather than bad moods. *Situational biases*, by contrast, are grounded in the subject's external environment, which can alter the salience of various affects or make certain ways of judging seem more appropriate. For example, subjects asked to rate their lives in the presence of a handicapped confederate who cannot hear their reports tend to report unusually high levels of life satisfaction.⁴⁶ Presumably, the sight of a handicapped person makes us prone to count our blessings. (While this study focused on life satisfaction, it is likely that reports of happiness would have yielded similar results.)

4. How extensive is AI?

How pervasive are these problems? To what extent is Glen's predicament mirrored in the general population? A sure answer to these questions obviously cannot be forthcoming given present measurement techniques. Indeed, I see no way of proving that some of these phenomena are even psychologically *possible*, much less pervasive: for the most part, the best evidence we've got relies either on retrospective assessments—"wow, I really did feel lousy"—or on indirect evidence, such as failures of self-reports to correlate with other measures of affect, such as ANS arousal. But retrospective assessments are themselves fallible, and failures to correlate with other quantities might reflect cases of affect (or related states) without the phenomenology.

So my arguments are not conclusive. But again, the primary aim is to establish as a serious possibility that substantial AI is widespread, to the extent that we should be concerned about our ability to assess how happy we are. (Such concern is consistent with the idea that people are mostly accurate about their happiness. Just because you aren't wrong about your happiness *most* of the time doesn't mean you shouldn't be worried that you're wrong *too much* of the time.) In any

⁴² Insofar as one's current experience is complex, with positive and negative affects, this bias might generate a form of present-AI as well.

⁴³ For discussion and references, see Headey and Wearing 1992, Cummins 1995, Diener and Diener 1996, p. 183, Argyle 1996, pp. 23-4, and Taylor and Brown 1988.

⁴⁴ Seidlitz 1993, cited in Diener and Diener 1996.

⁴⁵ For discussion and references, see Morris 1999, Schwarz and Strack 1999.

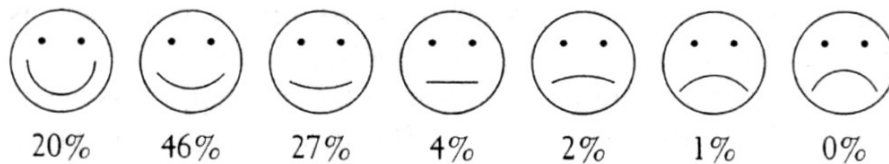
⁴⁶ Strack, Schwarz et al. 1990, cited in Schwarz and Strack 1991, Schwarz and Strack 1999.

event, I suspect that my contentions about AI will comport with the experience of many readers. Who among us hasn't at one time or another learned of their emotional conditions only through the testimony of observant friends and family members, or by attending to the physical manifestations of their emotional states, such as muscle tension or gastrointestinal symptoms best left to the imagination?⁴⁷ And who hasn't encountered blinkered souls who seemed spectacularly misinformed about their own emotional lives? In my own case, it seems that anyone wanting to know how happy I am would best consult my wife, as I am often the last to know. (I could, naturally, be wrong about this.)

Less anecdotally, some of the more striking results of subjective well-being research may well be due, in part, to AI. I have already mentioned that findings on hedonic adaptation may be somewhat exaggerated as a result of AI, specifically attentional adaptation, scaling adaptation, and affect-type biases. There is also some reason to think that self-reports of happiness, at least in the United States and similar countries, are systematically inflated, and that AI has something to do with this. For example, self-reported unhappiness tends to be very low in most surveys, typically less than ten percent and sometimes much less.⁴⁸ In the 1995 World Values Survey, for example, 94% of Americans reported being happy.⁴⁹ In one oft-cited study, only three percent offered negative responses (Figure 1). The percentage of subjects giving the most negative rating in this study was: *zero* (Andrews and Withey 1976).⁵⁰ About 93 percent gave favorable responses.

Figure 1. Taken from Myers 2000

Faces Scale: "Which Face Comes Closest to Expressing How You Feel About Your Life as a Whole?"



Note. The scale and data are from *Social Indicators of Well-Being: Americans' Perceptions of Life Quality* (p. 207 and p. 306), by F. M. Andrews and S. B. Withey, 1976, New York, Plenum. Copyright 1976 by Plenum. Reprinted with permission.

⁴⁷ Evidently, at least Schooler et al. have 2003, p. 45.

⁴⁸ Some of the higher "negative" response rates come from studies using a three-point scale where the most negative option is "not too happy" (vs. "very" or "pretty" happy). See, e.g., Gurin, Veroff et al. 1960, cited in Diener and Diener 1996, where 11% chose this option. But this option covers everyone from the slightly happy to the suicidal.

⁴⁹ Cited in Inglehart and Klingemann 2000. For reviews, see Myers 1992, Diener and Diener 1996, Myers 2000.

⁵⁰ While this study nominally assesses life satisfaction (which at any rate tends to yield similar, if somewhat less rosy, results), the response scale appears to measure happiness instead (Argyle 2002), and the results are more typical of happiness than life satisfaction studies. Some people blanch at the "smiley-face" format as silly or superficial, but bear in mind that, while the results are representative, the instrument itself is no longer popular, and there is a wide range of much more sophisticated measures in the literature.

These remarkable results are not anomalous: this study is often cited as a typical example of the results yielded by surveys.⁵¹ Such studies have been primarily responsible for the widespread conclusion among empirical investigators that most people are, contrary to popular belief, happy.

A majority of people may indeed be happy, but it is hard to escape the suspicion that the reported levels of happiness are a little overstated. Consider what we find using other methods. Perhaps the best source of data about people's affective lives is the experience sampling method (ESM), which relies not on global self-reports of happiness but on detailed on-line reports of subjects' experience roughly as it happens.⁵² One technique is to distribute Palm Pilots, which signal the subject several times a day to record their feelings at the time. This method is obviously subject to AI, but less so than global measures, particularly given that it doesn't rely on summary recollections of past experience. One important ESM study of 188 European subjects of varied backgrounds used reports gathered over a month through experience sampling (six reports per day per subject). The study found that feelings of sadness, fear, anger, or fatigue (a category that includes stress and exhaustion) were reported 34 percent of the time (Brandstatter 1991). If these reports are accurate, the average participant experienced these emotions more than *a third* of the time. Rather high, one might think. How happy can someone be who spends a third of her day being sad, angry, afraid, stressed, fatigued, or exhausted?⁵³

Consider also that the rate of self-reported unhappiness is in the neighborhood of many estimates of the rate of *depression*—the National Institute of Mental Health claims that about ten percent of Americans suffer from depression each year, with the average episode lasting around six months.⁵⁴ This figure may itself be considerably understated, as AI can infect depression measures as well. These instruments themselves rely on subjects' self-reports about various questions that are transparently aimed at assessing mental health, and hence are subject to both AI and a possible lack of candor. In a remarkable series of three studies by Shedler, Mayman, and Manis, such instruments were compared with the judgments of clinicians and physiological measures (heart rate and systolic blood pressure reactions to various psychological stressors such as counting backwards). The investigators found that a *majority* of those rated healthy by the scales were deemed to be distressed—as opposed to mentally healthy—in clinical evaluation.⁵⁵ In one of the studies, for instance, 16 of 74 subjects were rated genuinely healthy by both the Beck Depression Inventory (BDI) and clinical evaluation, whereas 21 of those rated healthy by the BDI were judged distressed by clinicians.⁵⁶ Referred to as “defensive deniers,” the latter group of individuals also exhibited much higher coronary reactivity in response to stressors than those rated healthy by the standard scales and clinicians, and as much as *double* the reactivity of those rated as distressed by both other measures. They also scored higher in verbal defensiveness during the psychological stress tests, and panels of undergraduates in one of the studies offered evaluations of the subjects that were similar to those of the clinicians. (The proffered explanation, corroborated by other stud-

⁵¹ Informal surveys given in my classes also yield similar results, with typically one or at most two students out of thirty reporting any level of unhappiness.

⁵² Stone and Shiffman 1999.

⁵³ Informal surveys conducted at the beginning of my intro ethics classes have found most students unwilling to ascribe happiness to individuals with this or a slightly less favorable proportion of positive to negative affect.

⁵⁴ 1999. Also, the one-year prevalence figures for major depressive episodes in two major U.S. studies are 6.5 percent (Epidemiologic Catchment Area study), and 10.1 percent (National Comorbidity Survey) (Satcher 1999). For a lower estimate, cf. Myers 2000, who cites a 2% rate of major depression and bipolar disorder at any given time (from the National Advisory Mental Health Council 1993).

⁵⁵ 1993. Clinical evaluations were based on subjects' descriptions of various personal recollections.

⁵⁶ The BDI correlated with self-reported happiness at -0.54 in the Fordyce study (1988).

ies, is that suppressing and concealing mental distress is costly, both psychologically and physically.⁵⁷)

It is certainly possible both that the clinicians overestimated distress and that some of the less extreme cases of distress might qualify as happy, or at least not unhappy.⁵⁸ (On the other hand, maybe some of those deemed healthy are unhappy.) But the corroboration from the physiological data, along with the highly negative appraisals for some of those classified as defensive deniers, strongly indicates that some, and possibly even most, of those rated as healthy by the self-report-based scales have serious deficiencies in their emotional conditions.

There is some plausibility, then, to the idea that depression measures understate the true rates of depression. And presumably some unhappy people aren't depressed. Let us not forget either the 40% of Americans who say their jobs are "very or extremely stressful,"⁵⁹ nor the alcoholics, those in abusive or crumbling marriages, the simply not-unhappy-but-not-happy-either, and so on. Perhaps people aren't quite as happy as they think they are.

My point is not to dispute the claim that most people are happy. Self-reports of happiness in countries like the United States are so relentlessly upbeat that they could be considerably inflated even with most people actually being happy. The point is just that we should take seriously the possibility that very many people are substantially mistaken about how happy they are. This in turn may reflect erroneous views about the nature of happiness or other causes besides AI. But we cannot, in light of what we have seen, dismiss the possibility that AI plays a major role in leading many of us astray about how happy we are. More importantly, we must allow that AI may be causing many of us to make serious errors about the quality of our lives.

5. Implications

5.1 For empirical research

I fear that the discussion of AI may lead some to conclude that empirical research on happiness is more or less worthless. Setting aside the fact that such research is highly diverse, some of it not even potentially subject to AI, this would be a serious mistake. Indeed, a principal aim of this paper is *reduce* skepticism about such research, as well as other uses of self-reports of affect in empirical studies, as in the search for neural correlates of consciousness.⁶⁰ For there is already considerable doubt among many investigators about the value of self-report measures, based largely (I believe) on an inchoate sense that such measures are subject to considerable error. The best reply to such worries, it seems to me, is to get clear on just where errors are liable to arise, replacing inchoate fears of an epistemic boogeyman with specific areas for concern. Those who want to challenge a given result will no longer be able simply to dismiss it. They will need to give an argument, showing where exactly we have reason to doubt the study's validity.

One reason for limiting skepticism is that self-reports can be informative even if profound AI were *ubiquitous*—indeed, even if no one had a clue how they felt. For self-reports can actually *gain* reliability from present-affect biases: such biases skew self-reports toward subjects' current affective states, and might do so even if the subject has no awareness of those states. People who feel good may simply be more likely to see their situations favorably, even if they don't realize

⁵⁷ For a discussion of "repressors" who seem clueless about high levels of anxiety, see Lambie and Marcel 2002.

⁵⁸ For instance, someone with a traumatic childhood may be happy but susceptible to extreme distress when made to reflect on the past (though some might question how happy such individuals really are).

⁵⁹ National Institute for Occupational Safety and Health (1999), citing a large study by Northwestern National Life.

⁶⁰ For a good collection of papers on the use of introspective evidence in cognitive science, see Jack and Roepstorff 2003, 2004. See also Jack and Shallice 2001 and Jack and Roepstorff 2002.

how they feel. In short, self-reports can convey information about affective states even when subjects don't know about them. Thus someone whose emotional condition is favorable will tend to offer more positive appraisals of her state even if she isn't aware of the relevant states. Averaged out over large samples, we should expect even the reports of utterly benighted individuals to correlate significantly with how happy they are.

Granted, other problems, including other sources of AI, might obscure the salutary effects of present-affect biases. But for these to cause worry we need some reason to think they are systematically skewing reports in a certain direction. For another point to bear in mind is that AI-related errors will often wash out over time, or over large samples. When you are assessing your own happiness, *any* errors are liable to lead you astray, so AI is much more worrisome. But for researchers studying large populations, many mistakes—e.g., a tendency to report being happier on sunny days—can be set aside as random “noise,” as they will tend to cancel each other out.

While self-reports may be useful guides to happiness even if we had no idea at all how we felt, this is obviously an extreme assumption. Even Glen, blinkered as he is, surely knows quite a lot. We can readily imagine that he normally knows when he is feeling particularly lousy, or playful, when he is angry or elated, and so forth. And that he typically can tell his better days from his worse ones. So while he fails to see a lot of important things about his emotional state, and mistakenly thinks himself happy when really unhappy, he could still tend *on average* to recognize it when he is happier, and when he is less happy. Thus an entire society of Glens, with an obscure and fragmentary awareness of their happiness, may well yield a good deal of useful information in surveys asking them how happy they are. And even this is a fairly pessimistic scenario: while we should consider the possibility that Glen's level of ignorance is the norm, the reality may not be quite that bad. We must also keep in mind that a variety of studies indicate that self-reports do correlate fairly well with happiness. For instance, there are substantial correlations between self-reports and other measures of affect, such as facial expressions, testimony of friends and family, and physiological measures.⁶¹

Consider the well-known findings that absolute levels of income and wealth tend to have weak effects on happiness, while unemployment has a very strong effect on it.⁶² These results, particularly the ones regarding money, have been widely replicated in a range of studies. So long as we have no reason to think that AI afflicts different groups of subjects differently in these cases, it should not worry us in relation to such claims. And it seems unlikely that the differences in reported happiness between employed and unemployed individuals are attributable solely to differences in the accuracy of their reports.⁶³

While these points offer little comfort to individuals trying to assess their personal welfare, they should temper skepticism about the science. Self-reports do give us useful information about people's happiness. But that doesn't mean they can't be terribly inaccurate in important ways. It doesn't mean that most people aren't, in stark contrast to self-reports, *unhappy*. Ironically, introspection may often be more useful in scientific contexts than for the conduct of our personal lives.

⁶¹ See, e.g., Diener 1994, Argyle 1996, Diener, Suh et al. 1997, Diener, Suh et al. 1999, Larsen and Fredrickson 1999, 2002, Larsen and Prizmic forthcoming. But note that self-reports could correlate perfectly with happiness yet still be systematically skewed—e.g., everyone might think themselves twice as happy as they are. So these points offer more support to correlational results than to claims about absolute levels of happiness. See Haybron forthcoming-a.

⁶² See, e.g., Diener and Biswas-Diener 2002, Diener and Seligman 2004, Argyle 2002, Frey and Stutzer 2002, and Layard 2005.

⁶³ On the other hand, consider the claim about money: perhaps financial gains lead to reduced stress, in ways that don't turn up well in self-reported happiness due to matters of elusiveness and affect-type biases. This seems unlikely to be a major effect, but is worth investigating.

5.2 For the study of consciousness

Ned Block has made an influential but controversial distinction between phenomenal consciousness (“p-consciousness”) and access consciousness (“a-consciousness”).⁶⁴ P-consciousness concerns the “what-it’s-like-ness” of our experience, its qualitative character, whereas a-consciousness concerns the availability of a state to serve as a premise in reasoning and in the rational control of thought and action—“awareness,” more or less. Block’s distinctions have met resistance in part because of doubts that p-consciousness and a-consciousness ever diverge.⁶⁵ And if there really are two types of consciousness here, it would be surprising if they *never* came apart: however strong the mechanism that links them, mechanisms break. For the most part, however, the examples offered have involved either science fiction cases or arcane perceptual experiments open to varying interpretations. Moreover, some, like David Rosenthal, argue that there can’t be phenomenality without the subject being able to tell, for how could there be anything it’s like *for the subject* if the subject doesn’t even know about it?⁶⁶

But if I am right about problems of elusiveness and adaptation, as well perhaps as variable attentiveness and discernment, there would seem to be a number of perfectly ordinary examples of p-consciousness without a-consciousness: affects like anxiety can impact the quality of our experience—there is something it is like *for us* to have them—even where we are wholly unaware of them; indeed, even where we *can’t* discern them. And if these problems arise at all—at least in non-disordered cases—then it is plausible that they are very common, given that my examples have not been particularly esoteric.⁶⁷

Now Block observes that a-consciousness does not *entail* introspectibility, or at least reportability, which makes sense given that young children and some animals may possess this sort of consciousness without either language or the ability to reflect on their mental states. So it could be argued that an indiscernible anxiety is a-conscious despite the subject’s inability to introspect it.⁶⁸ Perhaps, but one wonders in what sense this sort of anxiety is poised for “rational control” of anything—such anxiety may influence behavior and thought, but it appears to do so quite automatically and involuntarily, with no evident guidance from higher cognitive processes at all. Indeed, in situations requiring the heightened vigilance anxiety produces, there would be obvious disadvantages to burdening limited rational resources with the job of monitoring one’s affects and deciding what to make of them. Better to save those resources for dealing with the threats that triggered the anxiety.⁶⁹

⁶⁴ He has discussed this distinction in a number of places, but perhaps the most useful discussion appears in Block 1995. His more recent work emphasizes a-consciousness less and the notion of reflexive consciousness more, partly because of worries about whether a-consciousness is really a type of consciousness at all (see, e.g., Block 2001, 2003; for the worries, see Burge 1997). While I share these worries, a-consciousness appears to at least be a necessary condition for an important kind of consciousness, and it will be convenient to focus on this more familiar notion here. Moreover, it is harder and more interesting to establish phenomenality without access than without reflexivity.

⁶⁵ For the doubts see, among others, Church 1995, Chalmers 1997. Block himself seems to think that divergence is rare, and possibly nonexistent (Block 1995).

⁶⁶ See, e.g., Rosenthal 2002b, 2002a, and Rosenthal 2005, as well as Church 1995. I suspect this objection reflects an overly narrow view of the self and of subjectivity, but won’t pursue this question here.

⁶⁷ For further evidence, see a discussion of possible failures for “repressors” to discern anxiety in Lambie and Marcel 2002. For evidence that similar problems also occur for non-affective conscious states, see Schwitzgebel 2002. The limits to introspection described here appear to be quite general.

⁶⁸ I doubt Block himself would take this line, since his discussion of certain examples—like coming to notice, and *hence* becoming a-conscious of, a pneumatic drill that had been going on in the background—makes it plausible that he would count cases like this as p-conscious without a-consciousness.

⁶⁹ It is questionable whether such anxiety is globally accessible even to non-rational processes, so even a weakening of the notion of access may not help (see, e.g., Chalmers 1997).

It might alternatively be objected that even the most extreme cases of AI reflect, not p-without a-consciousness, but the simple misrepresentation of our affective states. We are aware of the anxiety, say, but not as such: we mistake it for something else. After all, if mood is so pervasive, then how could we *not* be aware of it at some level? Even Glen can easily attend to his mood—or so the objection might go—in which case he really is attending to his anxiety: his problem is that he doesn't recognize the anxiety for what it is. I will not dispute that some cases of AI can be explained this way. But it does not always seem like that: sometimes we seem to discover our anxiety, not by coming to a more accurate view of a feeling we'd already been aware of, but by *stumbling across* it, as something we had completely overlooked. Compare this with the experience of a hearing test, as when a tone begins so faintly that you aren't sure what you are hearing, or even if you are hearing anything; but as the volume grows, you recognize that you had indeed been hearing a certain sort of tone. Whereas a sudden change in your anxiety level can have the effect, not of making an existing perception of the affect clearer, but of making you aware of an affect you'd not been aware of at all. This objection arguably rests on a confusion about the nature of "mood," seeing an individual's mood as a unitary entity within the field of consciousness. If one is in an anxious mood, one need simply attend to one's mood to find it. But mood is not a singular entity like that: an individual's mood at any given time will normally involve a variety of states (one can, for instance, be both anxious and cheerful). And we can presumably attend to certain aspects of our moods without attending to others. Similarly, I am suggesting, we can be aware of certain aspects of our moods while missing others entirely.

While we might quarrel over the precise character of Block's distinctions, offhand it is not obvious how to make sense of these phenomena without invoking *something* like his distinctions. Dennett, for example, would likely try to explain problems like elusiveness in terms of richness of content and degree of influence: elusive affects have relatively impoverished contents, and/or lack influence.⁷⁰ This might explain how we can overlook, say, trivial pains. Yet a low-grade anxiety would seem to be anything *but* impoverished in content or lacking in influence. It infects virtually all aspects of our mental lives and has a tremendous impact on the quality of our experience. The phenomenology of anxiety may be difficult to pin down or articulate, but it is not lacking in richness. That's the *problem* with it.⁷¹ Similarly, attempts to explain putative cases of p- without a-consciousness as inattentive peripheral awareness appear unable to account for the rich phenomenality of unnoticed moods: if their phenomenality is purely the product of some sort of awareness, then their impact on our experience should be far less than it seems to be, just as more familiar examples of peripheral consciousness are exactly that: peripheral in phenomenal significance. Moreover, the phenomenality of moods does not seem to increase through attention as much as we would expect if it were purely a function of awareness.

None of this proves that Block's critics cannot accommodate the various forms of AI. For one thing, much of my evidence has itself been introspective. And perhaps the anxiety that eludes Glen's reflective efforts is nonetheless the object of a kind of low-level awareness, but one that need not involve accessibility to rational processes. (However, the preceding arguments suggest that this sort of awareness would have to play a special role in generating phenomenality, since the sort of awareness involved in attention seems not to contribute as much to the phenomenality of moods as we would expect otherwise.) But accommodation does seem required.

More importantly, it appears that conscious experience, specifically phenomenality, need not involve conscious awareness or introspectibility. This itself is a highly controversial claim, and might be granted even if we don't accept Block's distinctions or his arguments against functional-

⁷⁰ Dennett 1995. Rosenthal might raise a similar objection (2002b, p. 662).

⁷¹ Block has made some suggestive remarks along similar lines (Block 1995, p. 400).

ist accounts of mind. Thus experiments showing that certain mental states can occur outside of subjects' awareness may not involve states that are unconscious at all, for those states may be conscious but inaccessible.⁷² Similarly, the neural correlates for awareness may be quite different from those associated with phenomenality; to assume that the former provide the basis for conscious experience may be precipitate, and could lead to a faulty understanding of the mechanisms responsible for consciousness. It could also have unfortunate consequences for individuals with certain sorts of brain defects or damage. What if they are perfectly capable of suffering, yet we treat them as if they were not?

I hope it has also emerged that the study of consciousness can benefit from a greater focus on matters concerning well-being—which is, after all, the main reason we *care* about consciousness. The phenomena canvassed here are not just recondite perceptual quirks of purely theoretical interest; they threaten to undermine our authority about some of the most important things in our lives: the fundamental goodness or badness of our experience of life may be something that systematically eludes us. So much of our upbringing aims to facilitate the empathetic capacity to see what it is like to walk in other people's shoes. Yet we may often be distressingly blind about what it is like to walk in our own shoes.

5.3 For the study of well-being and politics

AI is one of several potential threats to our competence as seekers of happiness. Among other things, we may also be deficient at forecasting how choices will affect our future happiness,⁷³ at making choices that cohere with our considered judgments of value,⁷⁴ and at judging the value of psychological states commonly associated with happiness.⁷⁵ AI itself is a problem because it is hard to make good decisions concerning your well-being if you don't know how well you're doing.⁷⁶ If you think yourself far happier than you are, or badly overrate the quality of your experience, then you may become complacent and fail to address serious deficiencies in your quality of life. And if you think yourself unhappy when you're not—as restless youths may often do—then you may seek changes that undermine the happiness you didn't realize you had. In addition it is hard to learn from the past if you don't remember it correctly. For example, you might forget how miserable you were in a certain relationship or job and make a similar mistake in the future.

AI and other limitations of human psychology may thus lead people systematically to make serious mistakes in the pursuit of happiness. This raises troubling questions about the authority of the individual in matters of personal welfare, at least on the plausible assumption that happiness is highly important for well-being.⁷⁷ Such questions are obviously relevant for political thought. I will point to one of them here, but for obvious reasons will not try to settle any major

⁷² See, e.g., Berridge 1999, Berridge and Winkielman 2003, Winkielman and Berridge 2004. In one experiment, subliminal priming with happy or angry faces strongly affected subsequent desires for a novel beverage, though self-reports of affect were not influenced at all by the manipulation. This may well involve unconscious affect, but how confident about this can we be? Would you be indifferent to the prospect of an extended stint with the affective states induced by the angry faces, even if you were assured you wouldn't notice them, and even setting aside any behavioral consequences like those found in the study?

⁷³ See, e.g., Gilbert, Pinel et al. 2002, Wilson and Gilbert 2003, Loewenstein and Schkade 1999.

⁷⁴ See an important paper by Hsee, Zhang et al. 2003 as well as Tversky and Griffin 1991. Subjects in these studies tended to choose quantitatively (e.g., monetarily) superior options over options that would make them happier.

⁷⁵ See, e.g., Haybron forthcoming-a, forthcoming-b, where I argue that life satisfaction, with which happiness is often identified, is overrated.

⁷⁶ Schooler et al. put it more strongly: "Maximizing utility when one can't assess one's own level of utility is as futile as the task of a firm which tried to maximize its profits without being able to measure them" (2003, p. 42).

⁷⁷ I argue at length that it is in Haybron 2005. This should be plausible enough simply from the historical popularity of welfare hedonism. Hedonism may perhaps be false, but it isn't *crazy*.

issues. Consider that a deep faith in the ability of individuals effectively to seek their own good has provided an important justification for liberal restrictions on the state's role in promoting good lives. This strain of thought finds its classic expression in Mill's *On Liberty*, where he writes that "the strongest of all the arguments against the interference of the public with purely personal conduct, is that when it does interfere, the odds are that it interferes wrongly, and in the wrong place" (Mill 1991). Recall also the lines cited at the start of this paper. In essence, Mill argues that individuals tend to know how they are doing, and what's good for them, far better than anyone else does, and so societies should let individuals make their own decisions about how to live. Give people as much freedom to live as they wish, with as much scope for shaping their lives as they see fit, as possible.⁷⁸

And yet, if individuals are prone systematically to botch choices regarding their happiness, or even if this must be considered a serious possibility, then this aspect of liberal thought loses a good deal of its support, specifically the traditional consequentialist arguments like Mill's that favor it. We cannot simply assume a high level of prudential competence in the typical person. Nor can we assume, contra Mill, that governments won't often know better than individuals what's best for them, since some of our prudential shortcomings appear to be systematic. Thus policy-makers armed with knowledge of human psychological weaknesses might be able to shape social arrangements to compensate for them, in ways that will not always sit well with liberal sensibilities. One might object here that, as Mill claimed, individuals still tend to know their own affects better than anyone else does. But suppose, for the sake of argument, that most people mistakenly think themselves happy. Even if they are the best judges of their *specific* feelings, it may be that well-informed officials have a better grip on how the population feels, in general, than the individuals taken in aggregate do. So, for instance, state officials might know that the average person isn't happy, while the average person mistakenly believes herself happy.

Plainly, much more would need to be said actually to undermine consequentialist arguments for liberal strictures on state paternalism. Nor would the weakening or defeat of those arguments open the door for rampant government paternalism, since we could in any event have powerful reasons of autonomy for limiting state interventions in our lives.⁷⁹ My purpose here is just to show how AI and related psychological matters could impact political thought: we may find, perhaps among other things, that we need to rethink common doubts about the efficacy of state paternalism in making people happier.⁸⁰

6. Conclusion

The primary contention of this paper has been that the possibility of serious, widespread AI should be regarded as a live hypothesis, a prospect to be taken seriously. We cannot assume that people are reliable judges of their own affective states. In fact I suspect something stronger: that

⁷⁸ This way of putting it goes beyond Mill's main point about noninterference, but his discussion of individuality makes clear the importance he placed in having a wide range of options. At any rate, it well expresses a central element of modern liberal thought.

⁷⁹ For a critical overview of such arguments, see Sher 1997.

⁸⁰ For an interesting discussion, see Sunstein and Thaler 2003. Another theme in recent political theory has been a tendency to view the central goal of political decision-making as the promotion of individuals' effective freedoms or opportunities, either through the distribution of resources or through the direct promotion of individuals' capabilities or opportunities. Such views are not usually defended on the grounds that they are the best means of promoting well-being, but neither are they thought seriously to be *at odds* with people's welfare. Probably no one thinks their preferred ideal of the state would make people worse off. But the promotion of freedoms can compromise well-being insofar as people tend to make bad choices. While freedom matters well beyond its utility in bringing us happiness, we may need to attend more closely than we have to the human capacity to deal effectively with it.

pervasive AI is not just a possibility, but the reality. Most of my youth was spent shuttling between an undeveloped island, where my family spent part of the year, and a mainland home in Ohio. This afforded a Glen-like experience twice yearly, when we made the radical transition from the “mainland” mindset to the “island” mindset and back. The consensus among the island’s residents—at least so far as that anarchic bunch could register a consensus on anything—was that mainlanders were by and large a tense, small-souled, and thoroughly domesticated lot—not necessarily miserable, but not even close to the sunny results gleaned from self-reports of happiness. (Cf. Nietzsche on the “resentful man”: “his soul *squints*.”⁸¹) That at any rate was how we regarded the “touroids” who passed through the island like visitors from space, and for that matter how we regarded *ourselves* once the yoke of mainland respectability had reasserted itself (or at least until we had been away from the island long enough for the contrast to fade from memory). Naturally, these impressions might be somewhat romanticized. But I am hardly the first to record such thoughts.

There is a family I know—I will call them the Wilsons—whose members are quite amazingly loud. Wonderful people they are, but the din from their constant shouting, thumping, and crashing about is, for the unseasoned visitor, almost unbearable.⁸² Yet they seem to have no idea there’s anything at all unpleasant or odd about it, since it is perfectly normal for them. Those who know them see it differently: however hardened their sensibilities might have become, it’s almost certainly an unpleasant place for the family too. (It must be.) It is worth pondering whether mainstream American society might not be a little like the Wilsons: oblivious, and more or less inured to, a noisy, obnoxious, stressful, and spiritually deflating way of life.

7. References

- Andrews, F. M. and S. B. Withey (1976). *Social Indicators of Well-Being*. New York: Plenum Press.
- Argyle, M. (1996). “Subjective Well-Being.” In *In Pursuit of the Quality of Life*, ed. A. Offer. New York: Oxford University Press, pp. 18-45.
- (2002). *The Psychology of Happiness*. New York: Routledge.
- Armon-Jones, C. (1991). *Varieties of Affect*. Toronto: University of Toronto Press.
- Arntz, A. (1993). “Endorphins stimulate approach behavior, but do not reduce subjective fear. A pilot study.” *Behav Res Ther* 31(4), pp. 403-5.
- Bargh, J. A. and T. L. Chartrand (1999). “The Unbearable Automaticity of Being.” *American Psychologist* 54(7), pp. 462-479.
- Berridge, K. C. (1999). “Pleasure, pain, desire, and dread: Hidden core processes of emotion.” In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 525-57.
- Berridge, K. C. and P. Winkielman (2003). “What Is an Unconscious Emotion?” *Cognition & Emotion* 17, pp. 181-211.
- Block, N. (1995). “On a Confusion About A Function of Consciousness.” *Behavioral and Brain Sciences* 18, pp. 227-247.
- (2001). “Paradox and Cross-Purposes in Recent Work on Consciousness.” *Cognition* 79(1-2).
- (2003). “Consciousness, Philosophical Issues about.” In *The Encyclopedia of Cognitive Science*, ed. L. Nadel. London: Nature Publishing Group.
- Brandstatter, H. (1991). “Emotions in everyday life situations: Time sampling of subjective experience.” In *Subjective Well-Being: An Interdisciplinary Perspective*, eds. F. Strack, M. Argyle and N. Schwarz. New York: Pergamon, pp. 173-92.
- Burge, T. (1997). “Two Kinds of Consciousness.” In *The Nature of Consciousness*, eds. N. Block, O. Flanagan and G. Guzeldere. Cambridge, MA: MIT Press, pp. 427-33.

⁸¹ From the *Genealogy of Morals*. In Aristotle’s terms: the “*mikropsuchoi*,” perhaps? (Literally: “small-souled.”)

⁸² Some details have been changed to protect the innocent.

- Chalmers, D. (1997). "Availability: The Cognitive Basis of Experience?" In *The Nature of Consciousness*, eds. N. Block, O. Flanagan and G. Guzeldere. Cambridge, MA: MIT Press, pp. 421-4.
- Church, J. (1995). "Fallacies or Analyses?" *Behavioral and Brain Sciences* 18, pp. 227-247.
- Council, N. A. M. H. (1993). "Health care reform for Americans with severe mental illness: Report of the National Advisory Mental Health Council." *American Journal of Psychiatry* 150, pp. 1447-65.
- Csikszentmihalyi, M. (1990). *Flow: The Psychology of Optimum Experience*. New York: Harper & Row.
- (1999). "If we are so rich, why aren't we happy?" *American Psychologist* 54(10), pp. 821-827.
- Cummins, R. A. (1995). "On the trail of the gold standard for subjective well-being." *Social Indicators Research* 35, pp. 179-200.
- Dennett, D. (1991). *Consciousness Explained*. Boston: Little, Brown, & Co.
- (1995). "The Path Not Taken." *Behavioral and Brain Sciences* 18, pp. 252-3.
- Diener, E. (1994). "Assessing Subjective Well-Being: Progress and Opportunities." *Social Indicators Research* 31, pp. 103-157.
- Diener, E. and R. Biswas-Diener (2002). "Will Money Increase Subjective Well-Being?" *Social Indicators Research* 57, pp. 119-169.
- Diener, E. and C. Diener (1996). "Most People Are Happy." *Psychological Science* 7(3), pp. 181-185.
- Diener, E. and M. Seligman (2004). "Beyond Money: Toward an economy of well-being." *Psychological Science in the Public Interest* 5(1), pp. 1-31.
- Diener, E., E. Suh, et al. (1997). "Recent Studies on Subjective Well-Being." *Indian Journal of Clinical Psychology* 24, pp. 25-41.
- Diener, E. and E. M. Suh, eds. (2000). *Culture and Subjective Well-Being*. Cambridge, Mass.: MIT Press.
- Diener, E., E. M. Suh, et al. (1999). "Subjective Well-Being: Three Decades of Progress." *Psychological Bulletin* 125(2), pp. 276-302.
- Ekman, P. (1994). "All Emotions Are Basic." In *The Nature of Emotion*, eds. P. Ekman and R. J. Davidson. New York: Oxford, pp. 15-19.
- Ekman, P. and R. J. Davidson, eds. (1994). *The Nature of Emotion*. New York: Oxford University Press.
- Evans, G. W. and D. Johnson (2000). "Stress and Open-Office Noise." *Journal of Applied Psychology* 85(95), pp. 779-783.
- Fordyce, M. W. (1988). "A Review of Research on the Happiness Measures: A Sixty Second Index of Happiness and Mental Health." *Social Indicators Research* 20, pp. 355-82.
- Frederick, S. and G. Loewenstein (1999). "Hedonic Adaptation." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 302-29.
- Frey, B. S. and A. Stutzer (2002). *Happiness and Economics*. Princeton: Princeton.
- Frijda, N. H. (1993). "Moods, Emotion Episodes, and Emotions." In *Handbook of Emotions*, eds. M. Lewis and J. M. Haviland. New York: The Guilford Press, pp. 381-403.
- Gilbert, D. T., E. C. Pinel, et al. (1998). "Immune neglect: A source of durability bias in affective forecasting." *Journal of Personality and Social Psychology* 75, pp. 617-638.
- (2002). "Durability Bias in Affective Forecasting." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds. T. Gilovitch, D. Griffin and D. Kahneman. New York: Cambridge University Press, pp. 292-312.
- Glass, D. C., J. Singer, et al. (1977). "Behavioral and Physiological Effects of Uncontrollable Environmental Events." In *Perspectives on Environment and Behavior*, ed. D. Stokols. New York: Plenum.
- Glass, D. C. and J. E. Singer (1972). *Urban Stress: Experiments on Noise and Social Stressors*. New York: Academic Press.
- Griffiths, P. E. (1997). *What Emotions Really Are*. Chicago: University of Chicago Press.
- Gurin, G., J. Veroff, et al. (1960). *Americans view their mental health*. New York: Basic Books.
- Haybron, D. M. (2000). "Two Philosophical Problems in the Study of Happiness." *The Journal of Happiness Studies* 1(2), pp. 207-225.
- (2001a). "Happiness and Ethical Inquiry: An Essay in the Psychology of Well-Being." Ph.D. Dissertation, Rutgers University.
- (2001b). "Happiness and Pleasure." *Philosophy and Phenomenological Research* 62(3), pp. 501-528.
- (2003). "What Do We Want from a Theory of Happiness?" *Metaphilosophy* 34(3), pp. 305-329.
- (2005). "Happiness, the Self, and Human Flourishing." Submitted for publication.
- (forthcoming-a). "Life Satisfaction, Ethical Reflection and the Science of Happiness." *The Journal of Happiness Studies*.
- (forthcoming-b). "On Being Happy or Unhappy." *Philosophy and Phenomenological Research*.

- Headey, B. and A. Wearing (1989). "Personality, life events, and subjective well-being: Toward a dynamic equilibrium model." *Journal of Personality and Social Psychology* 57, pp. 731-9.
- (1992). *Understanding Happiness: A Theory of Subjective Well-Being*. Melbourne: Longman Cheshire.
- Health, T. N. I. o. M. (1999). "The Numbers Count: Mental Illness in America. (Washington, DC: The National Institute for Mental Health).
- Hsee, C. K., J. Zhang, et al. (2003). "Lay rationalism and inconsistency between predicted experience and decision." *Journal of Behavioral Decision Making* 16, pp. 257-272.
- Jack, A. and A. Roepstorff (2002). "Retrospection and cognitive brain mapping: from stimulus-response to script-report." *Trends in Cognitive Science* 6(8), pp. 333-339.
- , eds. (2003). *Trusting the Subject? Vol. 1*. Charlottesville, VA: Imprint Academic.
- , eds. (2004). *Trusting the Subject? Vol. 2*. Charlottesville, VA: Imprint Academic.
- Jack, A. and T. Shallice (2001). "Introspective physicalism as an approach to the science of consciousness." *Cognition* 79, pp. 161-196.
- Kahneman, D. (1999). "Objective Happiness." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation, pp. 3-25.
- (2000a). "Evaluation by moments: past and future." In *Choices, Values, and Frames*, eds. D. Kahneman and A. Tversky. New York: Cambridge, pp. 693-708.
- (2000b). "Experienced Utility and Objective Happiness: A Moment-Based Approach." In *Choices, Values, and Frames*, eds. D. Kahneman and A. Tversky. New York: Cambridge, pp. 673-692.
- Kahneman, D. and S. Frederick (2002). "Representativeness Revisited: Attribute Substitution in Intuitive Judgment." In *Heuristics and Biases: The Psychology of Intuitive Judgment*, eds. T. Gilovitch, D. Griffin and D. Kahneman. New York: Cambridge University Press, pp. 49-81.
- Kahneman, D., B. L. Fredrickson, et al. (1993). "When More Pain Is Preferred to Less: Adding a Better End." *Psychological Science* 4(6), pp. 401-405.
- Lambie, J. A. and A. Marcel (2002). "Consciousness and the varieties of emotion experience: A theoretical framework." *Psychological Review* 109, pp. 219-259.
- Larsen, R. J. and B. L. Fredrickson (1999). "Measurement Issues in Emotion Research." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 40-60.
- Larsen, R. J. and Z. Prizmic (forthcoming). "Multimethod Measurement of Emotion." In *Handbook of Psychological Measurement*, eds. M. Eid and E. Diener. Washington, DC: American Psychological Association.
- Layard, R. (2005). *Happiness: Lessons from a new science*. New York: Penguin.
- Lazarus, R. (1994). "The Stable and the Unstable in Emotions." In *The Nature of Emotion*, eds. P. Ekman and R. J. Davidson. New York: Oxford University Press, pp. 79-85.
- Levenson, R. W., P. Ekman, et al. (1992). "Emotion and autonomic nervous system activity in the Minangkabau of West Sumatra." *Journal of Personality and Social Psychology* 62, pp. 972-88.
- Loewenstein, G. and D. Schkade (1999). "Wouldn't It Be Nice? Predicting Future Feelings." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 85-105.
- Lormand, E. (1985). "Toward a Theory of Moods." *Philosophical Studies* 47, pp. 385-407.
- (1996). "Nonphenomenal Consciousness." *Nous* 30(2), pp. 242-261.
- Lykken, D. (1999). *Happiness*. New York: Golden Books.
- Lykken, D. and A. Tellegen (1996). "Happiness is a stochastic phenomenon." *Psychological Science* 7(3), pp. 186-9.
- Mill, J. S. (1991). "On Liberty." In *On Liberty and Other Essays*, ed. J. Gray. New York: Oxford University Press.
- Millgram, E. (2000). "What's the Use of Utility." *Philosophy and Public Affairs* 29(2), pp. 113-136.
- Morris, W. N. (1999). "The Mood System." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 169-189.
- Myers, D. G. (1992). *The Pursuit of Happiness: Who is Happy, and Why*. New York: William Morrow and Co.
- (2000). "The Funds, Friends, and Faith of Happy People." *American Psychologist* 55(1), pp. 56-67.
- Nagel, T. (1974). "What Is It Like to Be a Bat?" *Philosophical Review* LXXXIII(4), pp. 435-50.
- Nisbett, R. W. and T. D. Wilson (1978). "Telling more than we can know: verbal reports on mental processes." *Psychological Review* 84, pp. 231-59.
- Nozick, R. (1989). *The Examined Life*. New York: Simon and Schuster.
- Parkinson, B., P. Totterdell, et al. (1996). *Changing Moods: The Psychology of Mood and Mood Regulation*. New York: Addison Wesley.
- Prinz, J. J. (2004). *Gut Reactions*. New York: Oxford.

- Redelmeier, D. and D. Kahneman (1996). "Patients' memories of painful medical treatments: real-time and retrospective evaluations of two minimally invasive procedures." *Pain* 66, pp. 3-8.
- Redelmeier, D., J. Katz, et al. (2001). "Memories of colonoscopy: A randomized trial." Working paper.
- Rosenthal, D. M. (2002a). "Explaining Consciousness." In *Philosophy of Mind: Classical and contemporary readings*, ed. D. Chalmers. New York: Oxford, pp. 406-21.
- (2002b). "How Many Kinds of Consciousness?" *Consciousness and Cognition* 11, pp. 653-65.
- (2005). "Consciousness." In *The Encyclopedia of Philosophy, 2nd ed.*, ed. D. Borchert. New York: MacMillan.
- Satcher, D. (1999). "Mental Health: A Report of the Surgeon General. (Washington, DC: U.S. Public Health Service).
- Schooler, J. W., D. Ariely, et al. (2003). "The Pursuit and Assessment of Happiness Can Be Self-Defeating." In *The Psychology of Economic Decision*, eds. I. Brocas and J. Carillo. New York: Oxford.
- Schwarz, N. and F. Strack (1991). "Evaluating One's Life: A Judgment Model of Subjective Well-Being." In *Subjective Well-Being*, eds. F. Strack, M. Argyle and N. Schwarz. Elmsford, NY: Pergamon Press, pp. 27-47.
- (1999). "Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 61-84.
- Schwitzgebel, E. (2002). *The Unreliability of Naive Introspection*. Presented at the "Toward a Science of Consciousness" conference, University of Arizona.
- (2004). "Introspective training apprehensively defended: Reflections on Titchener's lab manual." In *Trusting the Subject? Vol. 2*, eds. A. Jack and A. Roepstorff. Charlottesville, VA: Imprint Academic. 2, pp. 58-76.
- Searle, J. (1990). "Who Is Computing with the Brain?" *Behavioral and Brain Sciences* 13(4), pp. 632-42.
- Seidlitz, L. (1993). "The organization and retrieval of valenced life events by happy and unhappy persons." University of Illinois, Urbana-Champaign.
- Shedler, J., M. Mayman, et al. (1993). "The Illusion of Mental Health." *American Psychologist* 48(11), pp. 1117-31.
- Sher, G. (1997). *Beyond Neutrality: Perfectionism and Politics*. New York: Cambridge University Press.
- Simons, D. J. and C. F. Chabris (1999). "Gorillas in our midst: sustained inattention blindness for dynamic events." *Perception* 28, pp. 1059-1074.
- Sizer, L. (2000). "Towards a Computational Theory of Mood." *British Journal of the Philosophy of Science* 51, pp. 743-769.
- Stone, A. A. and S. S. Shiffman (1999). "Ecological Momentary Assessment." In *Well-Being: The Foundations of Hedonic Psychology*, eds. D. Kahneman, E. Diener and N. Schwarz. New York: Russell Sage Foundation Press, pp. 26-39.
- Strack, F., N. Schwarz, et al. (1990). "The salience of comparison standards and the activation of social norms: consequences for judgments of happiness and their communication." *British Journal of Social Psychology* 29, pp. 303-14.
- Suh, E., E. Diener, et al. (1996). "Events and subjective well-being: only recent events matter." *Journal of Personality and Social Psychology* 70, pp. 1091-1102.
- (1998). "The shifting basis of life satisfaction judgments across cultures: Emotions versus norms." *Journal of Personality and Social Psychology* 74(2), pp. 482-93.
- Sumner, L. W. (1996). *Welfare, Happiness, and Ethics*. New York: Oxford.
- Sunstein, C. R. and R. H. Thaler (2003). "Libertarian Paternalism Is Not an Oxymoron." *University of Chicago Law Review* 70(4), pp. 1159-1202.
- Taylor, S. E. and J. D. Brown (1988). "Illusion and Well-Being: A Social-Psychological Perspective on Mental Health." *Psychological Bulletin* 103, pp. 193-210.
- Tversky, A. and D. Griffin (1991). "Endowment and Contrast in Judgments of Well-Being." In *Subjective Well-Being*, eds. F. Strack, M. Argyle and N. Schwarz. Elmsford, NY: Pergamon Press, pp. 101-118.
- Warshaw, M. G., G. L. Klerman, et al. (1991). "Are secular trends in major depression an artifact of recall?" *Journal of Psychiatric Research* 25, pp. 141-151.
- Wilson, T. D. (2002). *Strangers to Ourselves: Discovering the Adaptive Unconscious*. Cambridge, Mass.: Harvard University Press.
- Wilson, T. D. and D. T. Gilbert (2003). "Affective Forecasting." In *Advances in Experimental Social Psychology*, ed. M. Zanna. New York: Elsevier. 35, pp. 345-411.
- Winkielman, P. and K. C. Berridge (2004). "Unconscious Emotion." *Current Directions in Psychological Science* 13(3), pp. 120-123.
- Zajonc, R. (1980). "Feeling and thinking: preferences need no inferences." *American Psychologist* 35, pp. 151-75.
- (1994). "Evidence for Nonconscious Emotions." In *The Nature of Emotion*, eds. P. Ekman and R. J. Davidson. New York: Oxford University Press, pp. 293-7.