Hedonic Psychology, Political Theory, and Law: Is Welfarism Possible?

Mark Kelman
Stanford Law School

Follow this and additional works at: https://digitalcommons.law.buffalo.edu/buffalolawreview
Part of the Law Commons, and the Social Policy Commons

Recommended Citation
Available at: https://digitalcommons.law.buffalo.edu/buffalolawreview/vol52/iss1/3

This Article is brought to you for free and open access by the Law Journals at Digital Commons @ University at Buffalo School of Law. It has been accepted for inclusion in Buffalo Law Review by an authorized editor of Digital Commons @ University at Buffalo School of Law. For more information, please contact lawscholar@buffalo.edu.
Hedonic Psychology, Political Theory, and Law: Is Welfarism Possible?

MARK KELMAN†

INTRODUCTION

Legal policymakers and political philosophers confront, repeatedly and in familiar ways, questions about the propriety of using utilitarian metrics to evaluate discrete legal rules and social policy. Utilitarianism is doubtless the most familiar exemplar of consequentialist welfarism. Utilitarians are consequentialists in the sense that they believe that a policy or action should be evaluated solely in terms of the quality of the state of affairs that will arise if that policy or action is adopted. They are welfarists in the sense that the metric they use to judge the quality of the resulting state of affairs is the metric of “well-being”: the resulting state of affairs is better to the extent that we have increased the “sum” of how “well-off” the people affected by the policy or action are. A consequentialist need not try to maximize welfare. One could, for instance, evaluate one’s personal or political conduct with the goal of minimizing the number of rights violations that occur in the world. Most of the familiar attacks on utilitarianism center on the

† William Nelson Cromwell Professor of Law, Stanford Law School. The research was supported by the Stanford Legal Research Fund, made possible by a bequest from Ira S. Lillick and by gifts from others friends of the Stanford Law School. I am grateful for comments by Bernie Black, Paul Brest, Dick Craswell, Barbara Fried, Tom Grey, Daniel Kahneman, Norbert Schwarz, and Jeff Strnad, as well as participants at workshops at both Harvard and Stanford Law Schools. Errors, of course, are mine.
propriety of consequentialism: deontological theorists attack the notion that an action can be judged solely in terms of whether it improves some subsequent end-state. Thus, think of the familiar debate about whether an act that might be thought to violate X's rights is even permissible, let alone ethically mandatory, if it would “help” some individual Y or group of Ys more than it “hurt” X. This (important) dispute about consequentialism, though, is completely outside the scope of my inquiry.

What I want to focus on in this piece is welfarism. But it is not my goal in this piece to evaluate the traditional utilitarian claim, which has recently been restated with such vigor in the legal academy by Kaplow and Shavell, that we should judge the propriety of policies solely by reference to their welfare effects. Instead, I want to ask a more foundational question: to what extent is the concept of welfare meaningful, rather than problematic and elusive? Virtually all policymakers, whether they embrace or reject the notion that only shifts in welfare count in evaluating policy, believe that welfare effects are relevant, at least on some occasions. Similarly, in assessing our private actions, we would typically think it mattered, even if it were not invariably dispositive, whether we took action that made

1. See Louis Kaplow & Steven Shavell, Fairness versus Welfare (2002). It is important, though, even at this early stage, to draw a distinction between two sorts of facially political perfectionist challenges to welfarism. In a broad sense, perfectionists believe that it is the proper goal of political actors to insure that people live well-lived lives, or lives that express their “essential human natures,” rather than to facilitate their efforts to live whatever life increases their subjective well-being. I will address in detail only the second perfectionist argument in this piece. First, though, one variety of perfectionist might argue that welfarism itself must ultimately be seen to be perfectionist (despite the usual sense that perfectionism and welfarism are in stark opposition) simply because the third party observer’s decision to attend to welfare effects can only be justified by a higher order, foundational belief that people best fulfill their natures or live well-lived lives if they maximize their well-being. In that sense the question is whether the basic decision for a third party observer to care about welfare is just as perfectionist as the more overtly perfectionist third party decision to care that people’s lives are lived in a fashion that maximizes some moralistic virtue the observer cares for. I will instead be more interested in whether welfarists are covertly perfectionist in a second sense, though. Is it possible for the third party observer who claims to care about the well-being of those he observes simply to observe well-being from the vantage point of the subjects he observes or will he, inevitably, embody what I will call weak perfectionist preconceptions about the nature of the good life in attempting to elicit information about subjective well-being?
someone else “better off,” rather than “worse off,” or “better off to a greater, rather than to a lesser, extent.”

Welfarists are typically “subjective welfarists”—in the sense that they believe that a person’s welfare depends on her own attitudes about her experiences—and I will largely address the subjective welfarist tradition in this paper, for it is subjective welfarism that is associated with both liberal political theory and economics. But plainly “objective welfarism” is possible as well. One might believe that people themselves are well-off to the extent that they function in certain ways that are intrinsically good, independent of their private inclinations. What differentiates “objective welfarism” from conventional perfectionism is that what motivates the objective welfarist to proscribe a particular life canon is entirely his interest in the quality of life of the subjects themselves; he does not proscribe such a canon because the subjects are duty-bound to live a particular way or because the perfectionist herself feels bound to help realize a particular teleological end-state, in which people are realizing some underlying nature.

Subjective welfarism is justifiably tempting. It expresses a noble resistance to the idea that people are more or less the same, that what is good for Kate must be good for Jenny too, though they are distinct people who may have quite distinct grand life plans and quite distinct quirky little tastes. It also expresses the noble liberal idea that the question of whether Jenny’s life has been a good one must ultimately be answered (in some sense) from Jenny’s evaluative perspective if we are to respect Jenny as an autonomous person. There is, though, more ambiguity in the parenthetical phrase “in some sense” than subjective welfarists can comfortably accommodate, and that is, in essence, the subject of this paper. Jenny does not have a single evaluative perspective: she can reflect on her life in a host of ways, and our choice to treat one, rather than another, of these reflective evaluations as privileged in making judgments about how she has fared ultimately requires some sort of defense. No subjective welfarist has ever been up to the task of providing such a defense.

Broadly speaking, subjective welfarists have instead adopted two strategies, historically, to answer the question of whether (and to what degree) a person is better off with one state of affairs than another. The first, associated with early utilitarians like Jeremy Bentham, is to (try to) look
directly at sensation. From this perspective, a person is better off to the extent that she feels more pleasure and less pain. (Quite conventionally, I refer to this as hedonic utilitarianism). The second, associated with later Utilitarians and with modern economics, is to look at preferences: a person is better off if (a certain class of her) preferences are met or realized, rather than unmet. (Again following convention, I will refer to this as preference utilitarianism).

Each of these strategies is deeply flawed: my principal goal is to sharpen our understanding of the depth of the flaws. The most marked conceptual problem with hedonic utilitarianism is that it does not seem to take adequate account of subjectivity: there is no reason for any particular subject to care about whether she does or does not experience any particular set of sensations that we could describe as “pleasure” or “pain.” Later Utilitarians (and Economists) were driven to preference utilitarianism precisely for that reason: if, for instance, someone seeks (prefers) the glory of God to pleasure, why say she is better off if she instead experiences sensory pleasures?

But preference utilitarianism is just as problematic. At core, there is a fundamental conceptual problem in connecting the satisfaction of desire with experienced utility or well-being: desires are prospective and intentional while well-being is lived and experienced. Some preferences may be satisfied that we never experience (they may be fulfilled without our even knowing they have been fulfilled or after we are dead) and others may prove

---

2. The conceptual point is made quite clearly in many places: an atypically articulate one is James Griffin, Well-Being: Its Meaning, Measurement and Moral Importance 8 (1986) [hereinafter Well-Being].

3. For an excellent discussion of this basic conceptual point, see L. W. Sumner, Welfare, Happiness, and Ethics 122-32 (1996). Richard Brandt rejects preference-utilitarianism (or what he calls a desire theory of welfare) in favor of hedonic utilitarianism (he calls these happiness theories) for this reason as well. See Richard B. Brandt, A Theory of the Good and the Right 246-253 (1979).

4. For example, X strongly desires that his child grow up to be a lawyer; she does so long after he dies. Her becoming a lawyer fulfills his desire or preference, but under most plausible theories contributes nothing to his well-being. Obviously, the knowledge (or growing probabilistic sense) that she was on the lawyerly path might have given him satisfaction, just as he might have gotten pleasure from making efforts to realize his goal that she become a
disappointing. Not only is desire-satisfaction not a sufficient condition for improving well-being, it is not necessary either; just as we can be disappointed, so can we experience thoroughly unanticipated pleasures. While one might readily defend the proposition that there are some sound political theoretical reasons to respond to expressed preferences, whether or not such responsiveness increases subjective welfare, in this piece I endorse a variety of familiar arguments that the claim that it does inexorably increase welfare, non-tautologically understood, is a very weak one.

The most critical point I will try to elucidate is an obvious implication of the foundational observation that preferences are intrinsically prospective. Our preferences are merely predictions about the hedonic states that we will experience if certain end-states arise; like all predictions, they can be wrong. Thus, sophisticated preference utilitarians must try (and have invariably tried) to solve the problem of “error” by respecting only the sub-class of preferences most likely to be “correct”—those that are adequately informed and prudent. But what is critical to understand is that it is impossible to know when a desire is either prudent or adequately informed unless we know whether meeting it turned out to be hedonically satisfying, and we are drawn to preference utilitarianism precisely because we believed we could not observe such hedonic satisfaction directly. Thus, some psychologists (whom I will dub “the new hedonic psychologists” in this paper) want to “go back to Bentham” and measure hedonic states directly. The problem, though, is that the new hedonic psychologists are no more successful than Bentham was in avoiding the problem of inadequate subjectivity: there is no particular reason to believe that all people do or should value, above all, what the new hedonic psychologists describe as positive experience.

In Part I of this paper, I begin the detailed discussion by attempting to assess “hedonic utilitarianism.” My first main goal in that Part is to fortify a series of philosophical arguments that hedonic utilitarianism cannot cope with two deep foundational problems. First, it cannot cope with the argument that it is either covertly illiberally “perfectionist”
rather than appropriately subject-centered in assuming that there is a single variety of experience that everyone should seek or, more plausibly perhaps, that it is simply naive in ignoring the problem that people may be well-off whether or not they have any particular set of feelings. Second, it is beset by deep methodological problems: good and bad experiences are neither commensurable nor capable of being summed over a person's lifetime. I hope to do so, first, by highlighting the contributions hedonic psychologists have made to our understanding of the complexity of the notion of happiness. In this regard, it is my hope that I will be able to use hedonic psychology to supplement philosophy.

My second main goal in Part I is to analyze, critically, the efforts by the new hedonic psychologists to measure happiness by trying to sum subjects' spontaneous approach/avoid, continue/desist, good/bad reactions to their instant circumstances. These efforts were in part simply

5. I am interested in this paper, above all, in what could be described as the most "imperialistic" or all-encompassing claims about welfare that psychologists like Daniel Kahneman made in his earliest writings on hedonic psychology. To some considerable extent, Professor Kahneman has adopted a more moderate position in his later writings that I find less problematic, but in many ways less interesting. In ways I will try to hint at now, and supplement to some extent where most relevant, he now believes that the information about what he still refers to as "objective happiness" (essentially the sum of scaled momentary approach/avoid reactions) is not precisely information about "welfare" or human well-being, but merely one sort of information that a policymaker interested in welfare should find useful. Especially in this regard, see Daniel Kahneman, Experienced Utility and Objective Happiness: A Moment-Based Approach, in CHOICES, VALUES, AND FRAMES 673 (Daniel Kahneman & Amos Tversky eds., 2000) [hereinafter Kahneman, Experienced Utility]. It is important to contrast in this regard, "moderate statements" in the later work with earlier claims. For instance, compare (from the later work): "The concept of objective happiness is not intended to stand on its own, and is proposed only as a necessary element of a theory of human well-being . . . Objective happiness is only one constituent of the quality of human life, but it is a significant one." Id. at 683; and "The present analysis suggests that moment-based measures of the actual experience of consequences should be included in the assessments of outcomes and as one of the criteria for the quality of decisions. . . ." Id. at 690, with this earlier statement:

The present is fleeting, but memories and evaluations of the past endure and populate the mind. When they think about their lives, therefore, people have nothing to work with but memorized assessments and assessments of memories. The central claim of this chapter has been that the scientific study of enjoyment and suffering need not be subject to the same constraint. Real-time measures of experience can be obtained, stored without error, and aggregated to
responsive to perceived methodological flaws in the “old hedonic psychology”—which measured subjective welfare by eliciting responses to broad survey questions about whether subjects were content with their lives or not. We have very good reason to be suspicious of these conventional survey-based reports on subjective well-being—they really may pose questions that are simply cognitively intractable, even before we get to the question of whether they are asking questions about the “right sort” of happiness.

I argue, though, that the new hedonic psychology is philosophically naive: it is simply not the case that it would be appropriate to attempt to sum approach/avoid instantaneous reactions to life events and assume that those who approach the actual events in their life more often are “happier” than those who do not. This is true not only because there is no decent way of summing cardinally across incident reactions, and therefore accounting for the intensity of certain experiences; it is dubious as well, more profoundly, because people do not invariably (and probably should not) simply seek to maximize this sum. Their experience of any “life event chunk” is simply not appropriately conceived of as the sum of momentary pains and pleasures within those chunks, nor is their experience of their life as a whole inevitably simply the sum of the

yield a measure of objective well-being that is anchored in the reality of present experience, not the fallible reconstructions and evaluations of the past. . . . A combination of [measurement] methods will eventually be available to characterize the objective well-being of individuals and groups. . . . and to provide a criterion for the evaluation of economic and social policy.

Daniel Kahneman, Objective Happiness, in WELL-BEING: THE FOUNDATIONS OF HEDONIC PSYCHOLOGY 3, 22 (Daniel Kahneman et al. eds., 1999) [hereinafter WELL-BEING, FOUNDATIONS].

The newer claim is in some ways hard to argue with: so long as the policymaker “discounts” or “weighs” the information on objective happiness “appropriately,” it is hard to see why she would not want it. But it is also now much harder to fathom whether (and why) Kahneman believes that a welfarist should weigh this form of information a good deal (or barely at all) unless he explains why it “is” in some important sense information about welfare. Moreover, the later Kahneman is simply less clear than the earlier Kahneman about the normative status of memories of hedonically charged events: it is much clearer in the earlier pieces that memory is simply an imperfect recorder of information that could be stored (by an outside observer/recording mechanism) accurately, rather than an assembler of a distinct account of life experiences. Whether or not he in particular still believes that to be the case, it is an important view to confront.
hedonic valence of a host of “life chunk” experiences. Some subjects may privilege (certain forms of) reflective moments more than others; some may emphasize achievement, competence, and meaning more than sensation. I argue that we ultimately must be “weak perfectionists” in measuring happiness: we need not be strong perfectionists (holding views about the precise activities people should either like to do or value) but we must recognize that when we ask people to evaluate their experience from one of the many vantage points from which they might evaluate it, we are implicitly valuing certain attitudes about life more than others. Philosophical reflection will thus help us see that efforts to maintain a sharp separation between (subject-centered) welfarism and (evaluator-driven) perfectionism inexorably fail. The new hedonic psychologists may believe they have avoided the pitfalls of conventional Benthamite hedonic utilitarian thought—they do not assume that people value any particular sorts of sensations, but allow people simply to declare whether the state they are in is worth maintaining or not—but the belief that people who more frequently have the experience of wanting to approach, rather than avoid, present activity are better off is ultimately no more philosophically defensible than claims that they are better off experiencing certain “pleasurable” sensations. Thus, in a sense, while my first goal in Part I was (in essence) to use the findings of psychology to enrich philosophy, my second goal is to use the philosophical arguments to enrich psychology.

My third goal is to assess whether hedonic utilitarianism can be “rescued” by biology: perhaps there is but one form of happiness that we all (in fact) experience, and therefore can measure—the feeling that we get that drives us to make adaptive, pro-survival choices. I will argue that biology will not likely rescue us here, though my conclusions are especially cautious in this section, given what I am sure is my extremely modest comprehension of the debates among the relevant biologists. First, it is not clear that unhappiness is a biologically bad state (nor happiness a good one); thus, it would be misleading to identify “happy” people as those who were meeting “survival tasks.” Second, it is not at all obvious that

6. In a sense, my intuition is that some people might be drawn to using biology to finesse what I dubbed the first perfectionist challenge. See supra note
emotional states play any straightforward, single biological role if we were instead arguing simply that happiness (whether deeply functional or not) was a simple single state attached to the performance of a biological task.

In Part II, I discuss preference utilitarianism. As I noted, preference utilitarians can readily avoid the hedonic utilitarians' most transparent error—their tendency to impute a particular goal to people (the desire to maximize the excess of momentary "pleasure" over "pain"), whether they have got it or not. In this respect, they seem to embrace a philosophy more consistent with the most attractive aspect of subjective welfarism: its respect for individual diversity of taste. But preference utilitarians, in the end, cannot avoid the difficulties of hedonic utilitarianism; they have merely relocated them. Preferences are normatively meaningful only to the degree that they predict future hedonic reactions and predictions can be mistaken. Figuring out whether they are likely to be wrong reintroduces all the descriptive and normative difficulties of hedonic utilitarianism, innocently hidden in the concession that we must "correct" preferences to insure that they are made in procedurally appropriate ways.

What is critical to recognize is that we ultimately have no way of specifying the general procedural conditions under which these predictions are markedly more certain to be right. Maybe they are better when the party making them is more informed, both about the traits of the concrete goods or end-states she must choose among or more informed (in some nearly unfathomable sense) about herself—except when they are not. In this regard, I will once more try to use the psychological literature to enrich a point that some philosophers have repeatedly noted: more

1. In this regard, the argument is that third party evaluators have a non-controversial reason to be interested in maximizing the welfare of the subjects they observe: if those subjects are happy, they are fulfilling their biological destiny (engaging in pro-survival tasks.) Now one could argue that the intuition is philosophically naive—one should still have to justify why an observer ought to be happy to find out more people are engaged in survival-adaptive tasks. Instead, though, I will argue, more narrowly, that there is no biological reason to think that happier people are more likely to be meeting survival challenges, even if we could assume that a policymaker should feel that it is unambiguously good if people are meeting such challenges.

7. I will briefly discuss the fact that there are no welfarist reasons (though there may well be deontological reasons) to satisfy a preference simply because it is a preference.
information about the external world does not really give us more information about our preferences, though it may (but need not) help us meet our preferences, aptly defined. More information about whether our preferences are self-defeating is not precisely information at all: to say that preferences, in this sense, are wrong means they fail to meet some substantive end (other than the satisfaction of preferences). It almost surely means nothing more or less than that satisfying them fails to make us “happy” according to the hedonimeter that the preference utilitarians were trying to eschew. ⁸

Similarly, we cannot conclude that preferences are imprudently formed just because they form differently given distinct psychological background states. The philosophical point is that the only real ground for dubbing the choices made in one state imprudent is that they work out poorly, according to some non-choice-based criterion. In this regard, the psychology literature is especially informative in enriching the conventional philosophical arguments: even if “we” (or some of us) have some poorly defended, covert, substantively Puritanical notions that what psychologists describe as “cold” (low arousal) choices are more procedurally prudent than “hot” (high arousal) choices (regardless of which sort works out better hedonically, in the particular situation?), few of us have any intuitions about a host of other preference-influencing background conditions. For example, are more diversified one-shot choices more prudent than less diversified seriatim ones?

⁸ Ultimately, I will argue that the distinction between information about the external world and information about the self is less sharp than I imply here, but it is worth clarifying the initial intuitive distinction. If subject S wants to get from Palo Alto to San Jose as fast as possible, and chooses to take Highway 280 rather than 101, he might be mistaken if unaware of the traffic-stopping accident on 280. (Note that we have attributed a clear end to S, so that his mistake is plainly about the means of obtaining that end. Note, too, that a host of other information he might or might not have about 280 and 101—e.g., when each road was constructed, details about the road surfaces—would presumptively be of no moment: what we mean to say, though, in describing the increased information as unhelpful is that the information would not guide him to make a choice that led him to be happier.) If he does not “know,” though, that getting there more quickly will not make him happy, compared, say, to driving on a prettier road, he lacks some sort of self-knowledge. It is ambiguous what it would mean to be more “informed” about one’s own preferences, or that one would know whether one was “informed enough,” unless one were informed enough to make the choice that actually maximized hedonic utility.
We ultimately must resolve this question (if at all) not by reflecting on the dictates of procedural prudence but by sneaking a peek at which sort of choice made people substantively “happier” ex post. This is one more place where I hope that the philosophical debates will be clarified considerably by the findings of psychologists.

But the basic point of this paper is that full-blown subjective welfarism is not really possible. Preference utilitarianism both requires and defeats hedonic utilitarianism (it requires it to give content to the notion of informed and prudent preference and defeats it by highlighting the impossibility of canonizing any particular set of hedonic sensations). At the same time, there is no variant of hedonic utilitarianism that is either practically realizable (given the incommensurability of both good and bad states) or truly non-perfectionist (given the possibility that self-conception differs depending on the elicitation method we use to discover subjective states). All of us who rely to any extent on measuring the welfare of those affected by the choice of legal rules and policy initiatives—and a fortiori those who claim that lawyers and policymakers need attend to nothing but welfare—must really beware: as currently understood, the concept of subjective welfare is quite close to empty.

I. HEDONIC UTILITARIANISM, AS EMBODIED IN THE NEW HEDONIC PSYCHOLOGY

A. Introduction: The Development of the New Hedonic Psychology

Kahneman—representing what I dub the “new hedonic psychologists”—is quite explicit in his early writings on hedonic psychology that he intends to “go back to Bentham” by measuring utility directly, rather than

10. See id. Though hedonic utilitarians like Kahneman speak of returning to utilitarianism’s Benthamite roots, there are some indications, highlighted in an early twentieth century reflection by Wesley Mitchell, that Bentham himself was aware enough of the problems in “direct utilitarianism” (what I have been calling hedonic utilitarianism) that he often flirted with variants of preference utilitarianism. For a good discussion, see Wesley C. Mitchell, Bentham’s
Mitchell makes two especially salient points: First, Bentham recognized that one of the key components of the utility of an experience—its intensity—was not really directly measurable or discernible. Thus, there were only two ways of comparing two distinct pleasures (or pains): one was by making one pleasure or pain a part or ingredient of another. Thus, Bentham argues, “The only certain and universal means of making two lots of punishment perfectly commensurable is by making the lesser an ingredient in the composition of the greater. This may be done... [b]y adding to the lesser punishment another quantity of punishment....” Jeremy Bentham, *The Works of Jeremy Bentham* 92 (John Bowring ed., 1843). Note, though, that given the pain duration experiments, see infra text accompanying notes 36-38, even that will not really suffice: adding new pain to a fixed pain may not actually make the whole experience worse. (To the extent that people judge the misery of an experience by averaging the peak misery and the misery at the endpoint, adding a mildly bad experience to the end of a very bad experience may make it “seem”—be?—better than ending the bad experience entirely. Thus, patients may report less pain when the colonoscopy instrument is left in at the end of an otherwise-identical procedure even though the period in which it is left in is more unpleasant than the same period would be if it were taken out). It is transparent as well that this need not be true for conventional punishment—as well as pain infliction. There is no a priori reason to believe that longer terms of imprisonment are always hedonically worse than shorter ones, given a host of social and psychological responses to distinct terms (e.g., adaptation to prison; difficulty in reentering society; acceptance of more known hopeless situations rather than anxiety-provoking hope). See Mitchell, supra.

Bentham’s other suggestion, also anticipating preference utilitarianism, is to measure each pleasure or pain by imagining that we are in a barter situation or able to give each experience a money equivalent. Here is the relevant text, from Halevy’s translation of some of Bentham’s unpublished manuscripts: “If off two pleasures a man, knowing what they are, would as life enjoy the one as the other, they must be reputed equal... If then between two pleasures the one produced by the possession of money, the other not, a man had as life enjoy the one as the other, such pleasures are to reputed equal. But the pleasure produced by the possession of money is as the quantity of money that produces it: money is therefore the measure of this pleasure... Money is the instrument for measuring the quantity of pain or pleasure. Those who are not satisfied with the accuracy of this instrument must find out some other that shall be more accurate, or bid adieu to Politics and Morals.” Elie Halevy, *Radicalisme Philosophique* 410, 412, 414 (Mitchell trans., 1918). Like more recent preference utilitarians, including conventional economists, Bentham was aware that using money to measure pain and pleasure was especially problematic for those trying to do interpersonal comparisons because people with more money would pay more to enjoy smaller pleasures and pay more to avoid smaller pains. See Mitchell, supra.

Second, Mitchell argues that while Bentham may have professed to want to develop a Newtonian measurement system, it is more apt to compare his scientific efforts to Linnaeus than Newton: What Bentham ultimately does is to list and categorize multiple sources of pleasure and pain, but he cannot establish a system to calculate levels of either. Thus, Bentham notes (perhaps

---

relying on preference-satisfaction as a surrogate. His affirmative claim—which will help me frame the core arguments of this Part—is that happiness really is a measurable entity, and that we have developed both good subjective response technologies to measure a person's happiness over time, free from the recall and reporting distortions that have bedeviled "traditional hedonic" research and good "objective" (physiological, non-report-based, non-intentionally communicated) measures of happiness as well. This "better" response-based non-physiological measurement system would rely, first, on measuring subjects' simple good/bad, approach/avoid, persist/continue subjective reaction to large numbers of randomly selected moments in the subjects' lives rather than relying on their global satisfaction ratings of wider time periods and second, on monitoring certain involuntary physiological manifestations of what we have come to label as positive affect. For complex reasons I barely touch on, I rightly) that a person will devalue a pain more when it is longer-lasting, certain, or more pure, but in looking at any particular painful experience, this sort of general classificatory knowledge will be of very little use in supplying a cardinal utility scale for the experience. See Mitchell, supra.

11. The code name for this practice is "ecological momentary assessment." For a good review essay, see Arthur A. Stone, Saul S. Shiffman & Marten W. DeVries, Ecological Momentary Assessment, in WELL-BEING, FOUNDATIONS, supra note 5, at 26. It is a fascinating question in my mind—a "history of science" question that I do not feel I have the capacity to address in a serious fashion—whether one can (at least partly) attribute the recent "principled" theoretical attachment to this measurement mode to the invention of technological devices that made it feasible. Obviously, the invention of highly portable computers (like 'palm pilots') was a necessary condition to make the research agenda realistic; the thornier question is whether the availability of a new feasible measurement technology induces (to a significant extent) the production of an intellectual framework in which the "new measurement toy" is usable.

12. Basically, I think I believe that in order to use either the existing physiological measures of happiness or any conceivable measures (including ones that would depend on the direct observation of brain activity), we ultimately must end up concluding that particular physical changes correlate with properly-measured, self-reported hedonic states. (At core, there are two reasons for this: first, the physiological measures might, at best, produce ordinal measurements of whether the organism is more or less happy, but not cardinal measures that could be summarized across time for the organism; second, more conceptually and significantly, identifying that a particular brain wave pattern, for instance, is the "happiness" pattern requires interpretation of subjective data). As a result, the true barrier we face in measuring hedonic states is ultimately grounded in the need to insure that self-reports are
will focus on the subjective response technologies. There are a host of practical barriers to implementing the long-term program of the "new hedonic" psychologists, but I will not focus on these either.  

The interest in developing these techniques derives, at core, from negative reactions both to the preference utilitarianism that dominates modern economics that dominates modern economics on the one  

meaningful because if we find that physical changes correlate with states that are not themselves interpretable, we would have learned little of moment.

In the meantime though, there is a large cottage industry devoted to measuring or observing physical shifts that are thought to correspond to affective shifts. For instance, it appears possible that eye-blink rates (recorded by implanting miniature electrodes to measure activity in the orbicularis oculi muscle beneath the eye) are modulated by the affective valence of the experimental subject: the startle response seems to be lowest when subjects are (induced to be) "happiest" by being shown pleasant slides and highest when subjects had been exposed to "negative" slides. See Scott R. Vrana, Ellen L. Spence & Peter J. Lang, The Startle Probe Response: A New Measure of Emotion?, 97 J. ABNORMAL PSYCHOL. 487 (1988). We may also measure levels of cortisol secretions (generally measured from saliva samples); engage in Facial Action Coding through observation or electromyography—both of which, push comes to shove, are elaborate ways of quantifying how often we smile or furrow our brows; measure changes in autonomic activity, most particularly electrodermal activity; administer EEGs (and perhaps as cost goes down, PET scans or MRIs) that index patterns of anterior asymmetries thought to distinguish emotional states; use listeners or high-tech means to capture emotional state through vocal patterns, more particularly measuring both the level and changes in the level of what is known as fundamental frequency or pitch, the rate at which vocal folds vibrate and measure as well as the speed and intensity or loudness of speech. For a good review essay, see Randy J. Larsen & Barbara L. Frederickson, Measurement Issues in Emotion Research, in WELL-BEING, FOUNDATIONS, supra note 5, at 40. For discussions of the currently prevalent notion that people are "happy" when there is more electrical activity on the left, rather than right, hemisphere of the prefrontal cortex, see Richard J. Davidson, Affective Style and Affective Disorders: Perspectives from Affective Neuroscience, 12 COGNITION & EMOTION 307 (1998), and Andrew J. Tomarken et al., Individual Differences in Anterior Brain Asymmetry and Fundamental Dimensions of Emotion, 62 J. PERSONALITY & SOC. PSYCHOL. 676 (1992).

13. At core, the difficulty is that the techniques that the "new hedonics" people think are most accurate are difficult (and expensive) to use for large samples of people, yet we may often need to track happiness levels for such large samples to draw meaningful policy conclusions. As a result, the initial long-term "practical" goal of the new hedonic psychologists was to develop easily and cheaply mass-administered instruments to which responses correlate closely with the responses generated by what the researchers feel are the purer and more accurate measurement technologies that cannot be used by large numbers of subjects at a reasonable cost.
hand and the "old hedonic psychology" that dominated social psychology on the other.

It is not surprising that the "new hedonic psychologists" intuited the basic philosophical problem that bedevils preference utilitarianism: that preferences are only \textit{ex ante} judgments about the hedonic pay-off the subject \textit{expects} to derive from certain end states. But reflecting on the cognitive and affective difficulties that would make our predictions of the hedonic impact of different states of affairs undependable—and obviously, psychologists are prone to emphasize the degree to which assumptions of context-independent, rational judgment are problematic—makes it transparent that, in Kahneman's terms, "experienced" utility and both "predicted" and "decision" utility may diverge.\footnote{Kahneman puts it well: The weight that is assigned to the desirability of an outcome in the context of a decision is called its \textit{decision utility}. Decision utilities are inferred from choices and are used to explain choices. Much of the research on decision-making and on utility has been conducted in a rational and behavioristic tradition, which focuses on observable choices and shuns subjective notions such as experienced utility. It is implicitly assumed in this tradition that the experienced utility of outcomes can be inferred from their decision utility, because rational decision-makers surely know what they will like. Kahneman, \textit{Objective Happiness}, in \textit{WELL-BEING, FOUNDATIONS}, supra note 5, at 17.}

While I will not provide a fully detailed account of the negative reaction that the new hedonic psychologists have to the "old hedonic psychology," it is important to note a few basic points. Essentially, the critical claims that the "new hedonics" scholars emphasize are that the traditional survey-based findings rely on respondents answering a cognitively intractable question—how happy are you overall over some fairly extended period?\footnote{Typical questions are, "All things considered, how satisfied are you with your life as a whole these days?" Respondents were asked to use a 1 (dissatisfied) to 10 (satisfied) scale. Ed Diener & Eunkook M. Suh, \textit{National Differences in Subjective Well-Being}, in \textit{WELL-BEING, FOUNDATIONS}, supra note 5, at 435. \textit{See also} Ronald Inglehart & Jacques-Rene Rabier, \textit{Aspirations Adopt to Situations—But Why Are the Belgians so Much Happier than the French? A Cross-Cultural Analysis of the Subjective Quality of Life}, in \textit{RESEARCH ON THE QUALITY OF LIFE} 1, 7 (Frank M. Andrews ed., 1986) (questioning European Community respondents: "On the whole, are you very satisfied, fairly satisfied, not very satisfied or not at all satisfied with your life?"); David G. Myers & Ed Diener, \textit{The Pursuit of Happiness}, 274 Sci. Am. 70 , 71 (1996) (Survey questions include: "Do you strongly disagree, disagree, slightly disagree, neither agree nor disagree, slightly agree, agree or strongly agree: In most ways my life is close to my ideal. . . ").}—and that their
strategies to handle this intractable question reveal little about their "actual happiness." There are some good pre-theoretical reasons to distrust the conventional findings. Most particularly, critics of the conventional findings maintain that we observe low test-retest reliability of findings, results that are (unduly) sensitive to both short-term contextual influences that would seem too trivial to have a genuine impact on a stable sense of well-being, and high sensitivity of reported happiness to the construction of the survey instrument.

There are also good theoretical reasons to believe that the critics' skepticism is justified. Not only is it useful to consider this theory because it strongly bolsters the claim that skepticism about the conventional findings is justified, but, far more important for my ultimate purposes, reflecting on the theory will help us confront for the first time a number of vital questions to which I will need to refer throughout the piece: what might it mean to "be" happy, to "believe" oneself to be happy," to "misreport" whether one is happy, or "not to recall" whether one has

---

16. This is the dominant theme in Norbert Schwarz & Fritz Strack, Reports of Subjective Well-Being: Judgmental Processes and Their Methodological Implications, in WELL-BEING, FOUNDATIONS, supra note 5, at 61. Kahneman puts the point in a somewhat distinct way. In his view, traditional measures of happiness privilege memory over experience and memory is fallible. "Memory-based evaluations of experience and reports of current pleasure and pain are treated with equal respect in routine conversations, but the respect for memory is less deserved." Kahneman, Experienced Utility, supra note 5, at 676. "The memory-based and the moment-based views draw on different intuitions about what counts as real. There is an obvious sense in which present experience is real and memories are not. . . Because memories and stories of the past are all we ultimately get to keep, memories and stories often appear to be all that matters." Id. at 692.

17. Again, there is much more to the pre-theoretical debate that I do not delve into here, but it is worth noting the following. Question order effects on reported happiness are extreme: If one asks college students first, how happy they are and second, how many dates they have had in the past month, the correlation between the answers is but .12 but if one reverses the order of the questions, the correlation rises to .66. See Fritz Strack, Leonard L. Martin & Norbert Schwarz, Priming and Communication: Social Determinants of Information Use in Judgments of Life Satisfaction, 18 EUR. J. SOC. PSYCHOL. 429 (1988). Moreover, events as trivial as finding a dime may alter (what ought to be far less labile) feelings of well-being. See Norbert Schwarz et al., SOCCER, ROOMS, AND THE QUALITY OF YOUR LIFE: MOOD EFFECTS ON JUDGMENTS OF SATISFACTION WITH LIFE IN GENERAL AND WITH SPECIFIC DOMAINS, 17 EUR. J. SOC. PSYCHOL. 69 (1987).
The most trenchant theoretical key to understanding the contention that the broad survey questions about subjective well-being over substantial time periods are inevitably unhelpful is that such questions are cognitively intractable, so that those answering the

---

18. It might be helpful, in thinking about the distinction between “mistaken” and “distinct” modes of reporting, to think about two discrete problems in conventional hedonic survey reports. Experimental subjects are often asked to report, first, when certain hedonically-charged life events occurred, and second, how frequently such events occur; they are also asked the sorts of questions (about overall life-satisfaction or happiness) that I mentioned in the text. Answers to the first sorts of questions (dating events and noting frequencies) can be unambiguously mistaken (e.g., I take it, for instance, that events really occur on a particular date) and they frequently are: subjects “telescope” events (reporting that they occurred more recently than they did) and “estimate” rather than count the frequency of more commonplace events. For good review discussions of these problems in survey research generally, see Norbert Schwarz, Hans-J. Hippler & Elisabeth Noelle-Neumann, Retrospective Reports: The Impact of Response Formats, in AUTOBIOGRAPHICAL MEMORY AND THE VALIDITY OF RETROSPECTIVE REPORTS 187, 193 (Norbert Schwarz & Seymour Sudman eds., 1994) [hereinafter AUTOBIOGRAPHICAL MEMORY], and Norman M. Bradburn, Janellen Huttenlocher & Larry Hedges, Telescoping and Temporal Memory, in AUTOBIOGRAPHICAL MEMORY, supra, at 204.

When, though, subjects “add up” or average their pain experiences in distinct ways, or rate events with (relatively) happy endings as less stressful than they would have rated them at the time, it is not nearly so clear that they have made an “error” in reporting the “true” emotional valence/ongoing significance of the event. When the reports are heavily influenced by what seems like reporting artifacts—the manner in which responses are elicited, or the date on which responses are elicited, we have reason to be wary of the reports. (Thus, for instance, the fact that self-reports of the most recent week’s happiness level are unduly influenced by the affective valence of the most recent events makes weekly reports inappropriately sensitive to the precise moment at which the questionnaire is administered. See generally AUTOBIOGRAPHICAL MEMORY, supra.

But when events occur in the subject’s life that cause her to reconstrue more permanently the meaning of the event, the fact that the event is not averaged in to a lifetime summary of happiness as it was initially construed does not constitute an obvious “error” to believe that it does requires accepting what I will argue in the text are two quite controversial assumptions. First, one must believe that there are reasonable techniques of “summing” and “weighting” momentary hedonic experiences (e.g., that they are fully commensurable) and, second, and more important, that happiness is best thought of as the sum of momentary affective experiences, with little weight put on (some variety or other of) the hedonic status of an independent assessing “ego” judging both the prospective and retrospective attractiveness of relatively broad life-states (except to the degree that such assessments themselves provide momentary affective pleasure or pain).
questions must use simplifying strategies to come up with an answer. 19

There are a number of reasons for the intractability of questions that require integration of sensory perceptions over time, including perceptions of well-being. First, it is difficult to retrieve information from an extended time period. Though some relevant information is chronically accessible, much is not. As a result, respondents' replies may be grounded more in the availability of hedonically-charged information than its actual significance to a more informed and reflective subject. Survey respondents generally make judgments based on information that is accessible, generally because the relevant events are recent, 20 or the subject has been reminded of the information, perhaps by a prior question (deliberately or inadvertently included) on the same survey. 21

19. More generally, judgment tasks that require integration of sensory perceptions over time may be relatively intractable. Kahneman argues that most subjects could readily assess the brightness of an illuminated panel to which they are exposed, but if asked to assess the total, or average, brightness of the large number of panels they would have been exposed to in multiple trials, they would have no good way of doing so, and would more typically base their answer on the brightness of what they construed as some representative panel. See Kahneman, Objective Happiness, in WELL-BEING, FOUNDATIONS, supra note 5, at 3, 15. I am counterfactually assuming, in this piece, that subjects inevitably report the results of their hedonic introspection accurately. Obviously, though, the desire to present oneself in a socially acceptable fashion may cause subjects to edit their responses.

20. Even when subjects were asked to fill out relatively short-term assessments—to keep daily hedonic diaries—reports were unduly dominated by events at the end of the day. Events that occurred in the first three hours of the day had almost no impact on daily reports and end-of-day summaries were dominated by events that occurred in the most recent three-hour period, especially when reports were negative. See Stone et al., supra note 11, at 33.

21. For a general discussion of this phenomenon in survey response, see E. Tory Higgins, Knowledge Activation: Accessibility, Applicability, and Salience, in SOCIAL PSYCHOLOGY: HANDBOOK OF BASIC PRINCIPLES (E. Tory Higgins and Arie W. Kruglanski eds., 1996) [hereinafter SOCIAL PSYCHOLOGY]. For a general discussion of the fact that people typically stop searching for (inevitably hard-to-retrieve) information as soon as they get enough to deal with a task they are required to perform (like ascertaining one's hedonic state), see Galen V. Bodenhausen & Robert S. Wyer, Social Cognition and Social Reality: Information Acquisition and Use in the Laboratory and the Real World, in SOCIAL INFORMATION PROCESSING AND SURVEYING METHODOLOGY 6 (Hans-J. Hippler et al. eds., 1987).

The question order effects I mentioned, supra note 17, are generally assumed to arise because researchers can shift the information that subjects are most likely to retrieve by focusing attention on a certain sort of information
Second, though less interestingly, questions about global happiness asked in particular contexts are also difficult for subjects to interpret, given the existence of complex conversational norms that make it difficult for subjects to be certain they understand the sub-text, as well as text, of the questions they are asked to respond to. As a result, surveyors may (deliberately or not) influence whether the information that is retrieved is used or discarded simply by shifting conversational norms. For instance, given certain conversational norms about avoiding redundancy, if a subject is asked to comment on how satisfied he is with his marriage and then asked how happy he is with life, he may think he is being asked to disregard the impact that his marital satisfaction has on life satisfaction. Research suggests that he is markedly more prone to do so if the researcher does something as seemingly minor as telling the respondent that he is about to be asked two questions about his well-being.22

Third, and perhaps most interestingly, once information (about hedonically-charged incidents, about a hedonically-charged domain) is accessed and questions interpreted, the information must still be framed—essentially, either as representative of the subject’s current state or as belonging to some different state. If retrieved as representative of the current state, positive information inflates judgment of

---

22. When that lead in question is added, the correlation between life satisfaction and marital satisfaction drops from .67 to .18. See Schwarz et al., supra note 21, at 11. Similarly, the observed correlation between dates in the last month and life satisfaction drops back from .55 to .16. Id. at 5.
well-being (in standard parlance, it is "assimilated"); if thought to belong to some prior life, it will give rise to "contrast effects" (so that the retrieval of a good event will make one's current situation look bad in contrast). Thus, if respondents are asked to recall three recent events before answering a question about global happiness, they are more prone to report higher well-being levels if they recall positive events than if they recall negative events, but if asked to recall events that occurred more than five years ago, they report higher happiness levels if they remember negative events.

The "new hedonics literature" assumes that all of these sources of instability in research responses produce results that are merely "unreliable," in the sense that the respondents fail to report accurately on the presence or absence of an actual emotional state. They are surely right with respect to the first two of these sources of instability in survey responses—the memory/access problem and the conversational norm problem. I do not think the same can be said of the third source of instability the new hedonic

23. The notion that information could either be assimilated in this way or used to establish contrasts was first highlighted in Amos Tversky & Dale Griffin, Endowment and Contrast in Judgments of Well-Being, in SUBJECTIVE WELL-BEING: AN INTERDISCIPLINARY PERSPECTIVE 101 (Fritz Strack et al. eds., 1991).


Hedonically charged temporally distant memories may in some instances give rise to assimilation, rather than contrast effects: thus, whether retrieved incidents are recent or distant is a significant, but not determinative factor. It appears to be the case that if people recall a memory in vivid detail, it will first serve to alter their mood (e.g. a detailed recollection of a negative incident will induce temporary bad mood) and that their subjective well-being reports will then be mood congruent. For an experimental exploration of this complicating phenomenon, see id., experiments 2 and 3. (Subjects who gave long and vivid, rather than short, descriptions of negative events, as well as those who described how an event occurred rather than why, reported lower subjective well-being, while those who reported events in less emotionally involving fashion—short descriptions or why descriptions—displayed the usual "contrast" effect, e.g. the recollection of the distant negative event increased reported subjective well-being). The finding is replicated in a somewhat different form in Leslie F. Clark and James E. Collins, Remembering Old Flames: How the Past Affects Assessment of the Present, 19 PERSONALITY & SOC. PSYCHOL. BULL. 399 (1993).

25. For an excellent review essay, see Schwarz & Strack, supra note 16, at 61.
writers have identified (framing effects). The reason for this doubt goes to the central question I raise in this paper: Framing effects produce pure noise only if there is some objective state of subjective well-being apart from our mental construal of it—apart from the stories we tell ourselves about how happy we are. The new hedonics literature has staked its project on the assumption that there is some such objective state. In the version proffered by Kahneman, it is simply the sum of an aptly cardinalized measure of moments of instantaneous reactions to currently perceived events. If Kahneman's account is persuasive, of course, these framing effects are, in essence, the root of simple reporting errors because an event which has the same "actual" hedonic valence (e.g., one hour of intense physical pain) has a distinct impact on (reported) happiness (though the same impact on the sum of lifetime momentary pain) depending on whether it is assimilated (or used for contrast) at the moment of report. If, however, our

26. The fourth effect I will discuss—comparison of the self with a social reference group—is also at core a framing effect. See infra text accompanying notes 32-35. The fifth—duration neglect—is the most complex: looked at, as Kahneman does, as a product of memory deficit, it is a simple reporting error. But I will discuss whether it is appropriate to view duration neglect entirely as a reporting error, rather than a framed "construal" of event chunks.

27. This is most plainly true in his earlier writing, when he argued "objective happiness" was an appropriate all-inclusive measure of welfare, but it is even true to some degree given his more modest recent ambitions: momentary measures are still designed to measure something that should be described as an aspect of welfare.

28. Kahneman readily affirms the possibility that an ever-improving life—one in which contrast effects always lead to higher reported current happiness—is genuinely happier so long as the reason for that is that current—experiences are actually more pleasurable when contrasted with past events. But the only appropriate impact should be on current experience: past pains should not be "cancelled out" in reporting (because relegated to a contrast construct) nor should current experiences be reported as more pleasurable than they "actually are" simply because one responds to the cognitively difficult task of assessing actual hedonic state level by answering a simpler question, "am I better off now than I was during some past period I have readily recalled?" Kahneman treats the time path of experience as causing genuine hedonic changes (just as we will see later that he accepts that someone who recalls past pain, even incorrectly, might feel current pain merely from the recollection). See Kahneman, Experienced Utility, supra note 5, at 20 ("Without a doubt, the traveler who goes to a Kenya safari may continue to derive utility from that episode long after it ends, whether directly—by 'consuming' the memories in pleasant or unpleasant reminiscing—or, perhaps more importantly, by consuming the experience of the self as it has been altered by the event. However, the moment-based approach raises a question that should not be dismissed too lightly: How much time will
"authentic" hedonic state is a function, in some significant sense, of how we assess our lives on the whole, that affective state is not adequately captured by momentary responses to instantaneously perceptible activities: framing may be said to affect not merely reported well-being, but an integrated sense of well-being that is simply not readily captured by summing momentary assessments.

This conceptual point may perhaps be made most salient by considering that people may judge an event as positive not only if it is better than "past" events to which it is contrasted but if it is better than a readily-constructed counterfactual to which it is contrasted. Of course, there are situations in which this appears to lead to manipulable misreporting of what are (probably) underlyingly stable emotional states, grounded in surveyors' artifact, in just the way that "new hedonics" critics assume: After all, experimenters can (a) manipulate subjects' hedonic reports by directing them to construct either upward (more favorable than actual outcome) or downward (less favorable) hypotheticals or (b) can manipulate their tendencies to construct counterfactuals at all, knowing that if they do construct them, they might be more prone to feel regret over an outcome.29

---

29. See, in the first regard, Neal J. Roese, *The Functional Basis of Counterfactual Thinking*, 66 J. PERSONALITY & SOC. PSYCHOL. 805 (1994) (instructing participants first to record the details of a recent life event, then induce them to consider either upward or downward counterfactual alternatives to the event that they recorded; those induced to generate upward hypotheticals reported more negative affect). In the second regard, see, e.g., Dale T. Miller & Brian R. Taylor, *Counterfactual Thought, Regret, and Superstition: How to Avoid Kicking Yourself*, in *WHAT MIGHT HAVE BEEN* 305, 310-12 (Neal. J. Roese & James M. Olsen eds., 1995) (finding that experimental subjects playing simulated computer blackjack game are far more prone to experience affectively unpleasant regret when experimenters manipulatively make them believe that either picking a card, or staying, are "actions" rather than omissions; they believe they have taken action if they press "yes" when asked, "Do you want to hit?" or "Do you want to stand?" and overestimate their losses when they have taken action given the commonplace tendency to construct counterfactuals to actions one has taken in relationship to single, short-term events).
But what is one to make of the fact that even when these same framing effects arise spontaneously in the subject's life, they also appear to change reported subjective well-being substantially? With regard to the framing effect of hypothetical counterfactuals, consider the suggestive finding that winners of Olympic bronze medals display higher levels of happiness with their performance than silver medallists. Presumably the reason for this is that they imagine (as the relevant counterfactual) not winning a medal at all, and feel they have done better than that, while the second place finishers are disappointed because they imagine that they might have won the first place gold. In what sense might this represent a "reporting" error? In what sense could one say that they did not at least believe themselves happier? In what sense might we say that they were not "authentically" happier?

Fourth, survey responses might well be problematic because people may also judge their sense of well-being in comparison with others. This is, of course, yet another framing effect: just as a person may say she is happy because recalling a long-past sad event only reminds her that she is better off than that now, or may construe what might generally be thought an intrinsically unpleasant...
event so that it inflates subjective well-being because it appears better than constructed counterfactuals, so one may report, or construe, events more or less favorably depending upon whether one thinks they show that one is better off or worse off than those around her. This fact makes it possible, once more, for researchers to manipulate answers to questions about well-being by manipulating how subjects construct relevant comparison groups. For example, simply by implying norms of behavior in research instruments, people can be made to report higher levels of satisfaction or dissatisfaction with particular behavioral realms or their lives more generally. Once more, though,

---

32. For a good review of the literature on the degree to which both experimenters and “naturally existing conditions” may establish comparison norms, see Dale T. Miller & Deborah A. Prentice, The Construction of Norms and Standards, in SOCIAL Psychology, supra note 21, at 799.


33. See, e.g., Norbert Schwarz & Bettina Scheuring, Judgments of Relationship Satisfaction: Inter- and Intraindividual Comparisons as a Function of Questionnaire Structure, 18 EUR. J. SOC. PSYCHOL. 485, 488-89 (1988) (indicating that couples express lower levels of satisfaction with their sex lives if experimenters change the categorical boxes one can check for how often in a given time period they have sex so that the lowest frequency box one can check contains a larger number of sexual contacts in a given time period, as does the next lowest etc. because the subjects will believe they have less sexual contact than is typical if they are in the lower categories). See also Norbert Schwarz et al., Response Categories: Effects on Behavioral Reports and Comparative Judgments, 49 PUB. OPINION Q. 388 (1985) (finding that if respondents are asked to judge how satisfied they are with their leisure time activities, they will judge themselves to be more satisfied if given a questionnaire that elicits information about how much TV they watch each day where the lower range categories encompass a much larger amount of TV watching than when they are asked the question alongside a question about TV watching in which the lowest categorical amount of TV watching one can do
life itself creates the same framing effect spontaneously, in ways that may have persistent, rather than readily mutable, impact on reported subjective well-being. For example, students at a given level of performance have higher levels of self-esteem at low-quality schools and citizens in more egalitarian communities report higher levels of well-being than those with the same income in less egalitarian settings. Are we prepared to say that the "happier" citizens of more egalitarian communities are simply mis-reporting some more authentic, lower happiness levels? Can we say this even though these reports may not be so mutable that it seems reasonable to dismiss them as unreliable indicators of a persistent mental state we might care about?

Fifth, and again, quite significantly, new hedonic researchers note that existing reports are unreliable because subjects do not construct accounts of their well-being by summing the hedonic valence of all their experiences, but rather seem to think about accessible incidents that they treat as representative; thus, they typically neglect the duration of the pain and joy they have experienced. In judging how painful an event is, subjects'}

includes a far smaller amount per day since in the first condition, they will not believe they watch more TV than the average person).

34. See, e.g., Herbert W. Marsh & John W. Parker, Determinants of Student Self-Concept: Is It Better To Be a Relatively Large Fish in a Small Pond Even If You Don't Learn to Swim as Well?, 47 J. PERSONALITY & SOC. PSYCHOL. 213 (1984); Jerald G. Bachman & Patrick M. O'Malley, Self-concepts, Self-Esteem, and Educational Experience: The Frog Pond Revisited (Again), 50 J. PERSONALITY & SOC. PSYCHOL. 35 (1986). Note, though, that the enhanced self-esteem may be (at least in significant part) a function of how peers and teachers treat the relatively rarer successful students in poorer schools rather than a social comparison judgment.


36. For whatever it is worth, we seem to share this trait with rats; the little critters' level of aversion to a blinking light that has been paired with an electric shock that has been administered to them vary with the intensity or magnitude of the shock but not its duration. See O.H. Mowrer & L.N. Solomon, Contiguity vs. Drive-Reduction in Conditioned Fear: The Proximity and Abruptness of Drive Reduction, 67 AM. J. PSYCHOL. 15 (1954). Similarly, when scientists with more respect for rats' welfare (but equally little for their autonomy) administer pleasurable stimulation to the pleasure centers of rats' brains, the rats react only to the intensity of the stimulation, not its duration, in seeking out the stimulus. See Peter Shizgal, On the Neural Computation of Utility: Implications from Studies of Brain Stimulation Reward, in WELLBEING, FOUNDATIONS, supra note 5, at 500, 508-09. See also Terence A. Mark &
typical global retrospective evaluation conforms to what is called the Peak-End rule: they report the quality of the event as the arithmetic mean of the (presumably accessible) most painful moment and the (also easily recalled) final moment.

Thus, in evaluating how bad a particular medical procedure (colonoscopy) is, subjects will simply average the last pain they felt and the worst pain they felt. Thus, assume that Patient A experiences a peak pain of 8 (on a 10 point scale) and pain at the thirty minute mark of the procedure of 4, at which time the procedure ends, the instrument is removed and pain drops to 0. He will report a pain level of 6 for the procedure. Patient B has the same procedure as Patient A, and experiences the same level of pain at each moment, except that instead of removing the medical instrument at the half hour mark, it is left in for an additional fifteen minutes, producing a pain level of 2 (during a time, of course, at which A's pain level was 0). B will not only report a less painful procedure (a pain rating of 5, the mean of the peak of 8 and the end-point of 2), but will be more prone to choose to repeat the procedure when it is next medically indicated, even though he had experienced (on a moment-by-moment basis) just as much pain as A for a half hour, and more pain than A experienced for the next 15 minutes.\textsuperscript{37} Duration neglect is, under certain conditions, a phenomenon that is observed in the subjective evaluation of events.

\textsuperscript{37} See Donald A. Redelmeier & Daniel Kahneman, Patients' Memories of Painful Medical Treatments: Real Time and Retrospective Evaluations of Two Minimally Invasive Procedures, 66 PAIN 3 (1996). In a quite parallel experiment, subjects who immerse their hand in an unpleasantly cold (14 degrees Celsius) tub of water for sixty seconds report less pain if they then immerse their hand for an additional 30 second period in slightly warmer, but still aversively cold, (15 degree) water than if they remove their hand from the cold water at the end of the sixty second period. Their choices map up to their retrospective evaluation, as well. Sixty-five percent of participants choose to repeat the longer rather than the shorter trial. See Daniel Kahneman et al., When More Pain is Preferred to Less: Adding a Better End, 4 PSYCHOL. SCI. 401 (1993). For similar findings about duration neglect, see Carol Varcy & Daniel Kahneman, Experiences Extended Over Time: Evaluation of Moments and
conceptions, a simple reporting error; instead of accurately assessing one's actual hedonic state, one reports only that which one remembers most readily.\textsuperscript{38}

B. The Critique of Hedonic Utilitarianism

1. Introducing the Essential Problems: Commensurability, Scaling, and Unacknowledged Perfectionism. There are several issues raised in designing a hedonimeter. First, there is a practical problem: a hedonimeter can perfectly capture subjective well-being only if there is a single, scaled positive hedonic state (and a single, scaled negative one). That requires that all relevant positive (negative) emotions are fully commensurate with one another\textsuperscript{39} and that there are satisfactory ways to

---


38. I will discuss duration neglect at some length, see infra text accompanying notes 61-78, because I think it poses the most plausible (but still not fully convincing) exemplar for the "new hedonics" argument that reports about the hedonic quality of extended periods are less "accurate" than reports that sum moment-by-moment evaluations.

39. It is worth noting a very significant point here that Dick Craswell made in thinking about problems of commensurability generally: when looking at individuals, economists may duck the commensurability question by noting that individuals can make choices among option-sets, even when they believe in some way that the choices are not strongly commensurable. To track the terms I have been using, preference-utilitarians can largely duck the commensurability issue, simply by arguing that all options are commensurate (capable of meaningful comparison) in at least the sense that one can be preferred to another. (Naturally, this option is not available to hedonic utilitarians). But Craswell emphasizes that if we need to construct a social welfare function—requiring some capacity to weigh and measure how different people evaluate end states, incommensurability becomes a bigger problem. The reason in this regard, of course, is that there is no simple social summing/voting mechanism that resembles the choice of an individual. Thus, an individual may choose to sacrifice friendship and community for pay and we may think, as preference utilitarians, that is enough to endorse making material acquisition the normative end-state for her, even if she does not believe the choices are truly fully commensurable. But if we adopt a social policy that increases wealth for some while destroying old friendship and community bonds, for others, the fact that we do not have a common metric to deal with these gains and losses might be considerably more problematic because preference-summing techniques are inevitably controversial and hedonic utilitarian measures require direct commensurability and additivity. \textit{See} Richard Craswell, \textit{Incommensurability, Welfare Economics, and the Law}, 146 U. PA. L. REV. 1419, 1447-50 (1998).
measure (in a fashion that permits summing, through explicit cardinalization\textsuperscript{40} or some technique for transforming ordinal rankings into a "welfare function") how positive or negative any particular hedonically-charged state is.\textsuperscript{41}

At the same time, the basic conceptual disquiet with direct hedonic utilitarianism is easy enough to explain. There is no reason to believe that people do (or should) value any particular concrete end-state. If we give any meaningful content to "pleasure" or "pain," then there is no reason to assume that a rational agent should maximize net pleasure, rather than a host of alternative emotional end-states. People might rationally choose instead to maximize the degree to which others admire them, the degree to which they meet perceived duties, the degree to which they recall events in their life vividly (even if the events are all vividly unpleasant). Now, it is possible to say that anything people choose to seek should be defined as pleasurable (so that we avoid the seeming narrowness and implicit

\textsuperscript{40} I should make reference in this regard to the two most mainstream techniques for deriving "cardinalized" utility judgments from ordinal pair-wise comparisons: one involves the observation of probabilistic choices and the other involves successions of pair-wise choices in which one member of the pair is more plainly cardinal. Thus, one possible way to say that we value experience X at least twice as much as experience Y is to say that we would choose a 50% chance of X over a 100% chance of Y. (And we could "perfect" our cardinal measurement scheme by noting, for instance, that we would choose certain Y over a 33% chance of X). We might also argue that we value X twice as much as Y if we are willing to work twice as many hours to get it (at least if we assumed the disutility of work was related to its duration in a linear fashion). For a good discussion of these standard methods, see Griffin, supra note 2, at 101 (1986).

\textsuperscript{41} I suspect that those interested in exploring the full range of the ultimate policy-relevance of hedonic psychology must also at some point attend to the question of whether we can properly measure any given hedonic state, even assuming there are a range of such states that are impossible to compare with one another. And, of course, one must then deal with a significant ancillary issue: assuming one could solve the measurement problem for a single state, would good information about that particular state contribute to policy discourse? Here is a useful mini-hypothetical, though, that might clarify what the concern is in this regard. We may believe that in designing a state policy or set of—formalized or informal—professional norms about pain medication administration, it would help us to know just how much purely physical pain people felt at certain points of time. We might want to know this information even if we thought the measures of such pain were in significant ways incommensurable with other relevant hedonic information (including hedonic information that is transparently relevant when thinking about pain-reduction treatments): how do we add up pain and awareness, pain and loss of control, pain and shorter life?
perfectionism of a welfarism that is, at its deep core, plainly designed to put few constraints on the sway of individual choice), but if we simply define the things people seek as the things that give them pleasure, we seemingly might as well move away from direct hedonic utilitarian thought, with its implicit psychology of motivation. In this sense, preference utilitarians cannot be satisfied that the new hedonic psychologists do not explicitly extol any particular sensations (like “pleasure” or “pain”) in the fashion that conventional hedonic utilitarians may have, but merely report on whether a person responds “favorably” to his situation. There is no rate of such “favorable response” that is unambiguously optimal for all subjects.

I make observations on the intersection between this conceptual problem and the first practical requirement in hedonimeter design in both subsections (b) and (c). There, I will explore the implausibility of the necessary assumptions that there either be a single end state to which we aspire or a set of easily commensurable end states. I then return, in (d), to one effort to defend the proposition that at core, “positive” and “negative” affect have a single meaning, dictated by the purported (single, simple) biological functions of each. I will largely ignore the very difficult issues of scaling raised by the second practical requirement.

42. Perhaps surprisingly, making this move may not take us all the way to preference utilitarianism because a true preference utilitarian must account for the meta-taste to meet, or not meet, one's discrete preferences. See infra text accompanying note 46.

43. To the degree that the new hedonic psychologists ultimately look to measure happiness by measuring physical states directly, rather than reporting intentionally communicated information, see supra text accompanying note 12, then the accusation that they are subject to the critique that hedonic utilitarianism is inadequately subjective obviously gains even greater force. The physical measures really are quite plainly either particular sensations or proxies for such sensations.

44. I do want to note a few points relevant to the possibility of scaling that I hope I (or others interested in exploring the relevance of hedonic psychology to political theory and law) must eventually explore at greater length. First, there is an important and (possibly) counterintuitive finding to report, lying at the border of the scaling issue and a huge issue to which I return soon, whether the degree to which we are happy is always best judged from the perspective of an integrated Ego judging her life rather than the perspective of a third party attempting to integrate momentary experiences in a fundamentally mathematical way. If one takes the global survey (how happy are you?) questions at all seriously, it would be interesting to find out (as researchers
2. Commensurability of End States or the Selection of a Single State. Many of the neo-utilitarian philosophers' attacks on Bentham's "direct" hedonic utilitarianism focused on the observation that there was no reason to believe that people should seek or value any single end-state, nor should one believe, as Bentham did, that as a matter of description, the single-minded pursuit of (something we might label) pleasure (or the "excess" of pleasure over pain) in fact motivated actors. It is not even inevitably or logically the case that we seek a single end if

have) that the intensity of positive experiences does not predict overall happiness, what matters is the sheer number of good experiences. (This implies, in some versions, that we should not scale distinct good experiences at all if we were attempting to measure hedonic states over time—which seems worse than counterintuitive—or, even more counter-intuitively, that we should scale but count seemingly hyper-pleasurable experiences as less valuable than milder positive experiences). See, e.g., Ed Diener, Ed Sandvik & William Pavot, Happiness is the Frequency, Not the Intensity, of Positive versus Negative Affect, in SUBJECTIVE WELL-BEING: AN INTERDISCIPLINARY PERSPECTIVE, supra 23, at 119 (suggesting some plausible explanations of why it might be true that hyper-pleasurable experiences are counter-hedonic: for instance, really good experiences give rise to unfortunate contrast effects thereafter that mildly pleasurable ones might not; hence, the possible implication that we should assign a lower positive number to very pleasurable experiences).

My point, though, is not whether the finding is right; it is that it is plausible and that its plausibility makes it hard to imagine what sort of summing scheme we could use for pleasurable and unpleasant moments. The amount of money you have accumulated over time is transparently the sum of what you have taken in minus what you have expended and given away—imagine thinking someone is richer than his neighbor simply because he has been paid more often and gone to the store to shop less frequently. It is not so clear that the amount of happiness in your life is in some parallel way the sum of the amount of happiness you have experienced.

45. At core, "preference utilitarianism"—the dominant form of modern utilitarianism—is grounded to a significant extent in the view that rational beings need not endorse Bentham's notion that the avoidance of pain and the pursuit of pleasure were their proper despotic masters. See, e.g., John C. Harsanyi, Morality and the Theory of Rational Behavior, in UTILITARIANISM AND BEYOND 39 (Amartya Sen & Bernard Williams eds., 1982). Harsanyi states: [H]edonistic ... utilitarianism... presupposes a now completely outdated hedonistic psychology. It is by no means obvious that all we do we do only in order to attain pleasure and avoid pain.... Even if I want to accomplish something for myself, it is by no means self-evident that my main purpose is to produce some feelings of pleasure in myself, and it is not the accomplishment of some objective condition, such as having a good job, solving a problem, or winning a game etc.... [T]here is no reason why any theory of morality should try to prejudge
we decide to define that end as inclusively (and near-tautologically) as possible (e.g., we seek the satisfaction of our desires, whatever they may be). We might, for example, have the meta-desire *not* to have each desire satisfied, so that it would be wrong to argue that our goal was the satisfaction of each of our desires. Thus, a true preference utilitarian might note that while subject S ordinarily sought some list of positive end states A, B, C... Z, and preferred each to some ordinal-pair list of not-A, not-B... not-Z, he still did not prefer to have all his desires met. The true preference utilitarian notes not just that A may not be “more pleasurable” (in any narrow sense of what the word “pleasurable” might mean) than not-A, but that the subject need not choose to maximize the occurrence of favorable experiences, at all, however he defines favorable experiences.

For instance, many of us might think that being drugged into a constant state of satisfied euphoria was a bad outcome (implying we value something like “self-actualization” or “control” as well as experienced pleasure), even if we genuinely thought the drug did

---

the issue whether people are always after pleasure or whether they also have other objectives.

*Id.* at 54.

46. For a fuller argument to this effect, see Anthony Kenny, *Happiness, in LXVI PROCEEDINGS OF THE ARISTOTELIAN SOCIETY* 93, 95. Kenny notes that this meta-desire is no more problematic or internally contradictory than the belief that not all of our beliefs are true (though we neither know which beliefs to discard nor which desires should go unsatisfied). “Modesty seems to demand that... we should believe that some of our beliefs are false... Similarly, patience seems to demand... that we should be willing that some of our desires should be dissatisfied.” *Id.* at 95.

47. The question of whether we value the pursuit of a chosen life plan matters a good deal even when we are not evaluating the lives of those drugged into a passive state of bliss; it matters when we try to understand conventional welfare economics propositions. For a standard statement of the view that we need to look at the process by which we attain “happy end states,” see Amartya Sen, *Capability and Well-Being, in THE QUALITY OF LIFE* 30, 39 (Martha C. Nussbaum & Amartya Sen eds., 1993). Sen states:

[F]reedom may have intrinsic importance for the person’s well-being achievement... not just because more freedom may make better alternatives available. This view is contrary to the one typically assumed in standard consumer theory, in which the contribution of a set of feasible choices is judged exclusively by the value of the best element available. Even the removal of all the elements of a feasible set... other than the chosen best element is seen, in that theory, as no
induce "pleasure."\textsuperscript{48} We may also think it is of little or no consequence if we discover that we have been "happy" because we have lived under such rigidly oppressive circumstances that we have had few wishes or aspirations (implying that we value the development of challenging aspirations as well as satisfaction).\textsuperscript{49} Those who have strong senses of duty may lead just the life they have sought without attending at all to their own pleasure and pain

real loss, since the freedom to choose does not, in this view, matter in itself.

\textit{Id.} at 39. From the vantage point of more conventionally utilitarian economists, the value of "freedom" (e.g. of larger option sets) is simply a function of our uncertainty about our future preferences. For the standard economic formulation, see David M. Kreps, \textit{A Representation Theorem for 'Preference for Flexibility,'} \textit{47} \textit{ECONOMETRICA} 565 (1979).

I suspect that the best hedonic welfarists' response to the 'freedom' issue might go something like this: presumably, freedom has no intrinsic meaning, it is simply a contingent empirical question whether people like it (for Benthamite direct utilitarians) or choose it (for Hare's preference utilitarians). Surely, it is possible that Agent A would prefer the state, "I chose X over Y from the option set \{X,Y\}" to the state, "I was given X" and it is also possible that the process of choosing makes him giddy (in the properly reported subjective well-being or the startle probe sense); but it seems plausible as well that he would choose not to make choices or that the choice-making process leads to anxiety, declining self-reported subjective well-being, and all the bad physiological signs. At that point, it strikes me that when Sen speaks of freedom's "intrinsic importance," he either has a non-welfarist, perfectionist view of freedom—you \textit{should} want it whether you want it or not—or, more weakly, that he claims that, at least when you do not overtly devalue it, the way in which you like it is not especially commensurate with the gains from the consumption of X. I think it is the second point he is really trying to make, but I am not really sure that reading is persuasive. \textit{See Sen, Capability and Well-Being, supra.}

48. This is one conventional reading of the message of anti-Utopian works. \textit{See generally ALDOUS HUXLEY, BRAVE NEW WORLD} (1911).

49. For more in-depth discussions see generally \textit{PARTHA DASGUPTA, AN INQUIRY INTO WELL-BEING AND DESTITUTION} (1993); Amartya Sen, \textit{Capability and Well-Being, supra} note 46. There are two distinct perfectionist reactions to this sort of observation. In the one (more subject-focused) variant that is my main focus, we discount momentary reports of happiness because we believe the subject herself devalues this sort of happiness and would in some sense "enjoy" another. In another view, the third party observer disclaims interest in aspiration-impoverished happiness, believing that even if the subject never experiences any discontent with her life-state, she has still not lived as good or fulfilled a life as she would have had she had more opportunities to develop higher aspirations. This second view is more associated with the view that subjective welfarism is too thin a theory to motivate all social and political action, rather than a theory associated with the view that it is incoherent.
IS WELFARISM POSSIBLE?

The basic conceptual problem with the simplest form of hedonism is that it is not, when pressed, really a subjective welfarist theory at all. If a hedonist assumes there is some particular set of sensations that everyone should seek, he is overriding the authority of the subject to determine whether those sensations are really most significant to him. And even if the hedonist backs off and says that the sensations that make people “happiest” are, by definitional fiat, those that they most “enjoy” (thereby trying to blur the line between hedonic and preference utilitarianism), he is still subject to the challenge that he could not readily explain why some value only veridical or authentic, rather than illusory, experiences that could generate the same sensations. The choice is sensible only for those with some meta-desire that does not value just sensations themselves, but also conditions that do not have a sensory expression. Thus, some people may prefer actual over illusory experiences because they have the meta-desire to engage their active (doing) faculties or may have the meta-desire not to engage in self-delusion more generally in evaluating how their lives have gone.

Second, the category of “pleasures” surely includes a wide variety of sensations that are conceptually difficult to compare. A careful reading of the hedonic psychology literature bolsters this philosophical observation substantially. How do we compare (on any single scale) satisfactions gained from (what some might term) “comfort”

---

50. For a good discussion of why one cannot profitably argue that anything that people in fact seek describes what makes them happy, see Kenny, supra note 46, at 99.

51. For an excellent discussion of this conceptual point, see Sumner, supra note 3, at 92-98.

52. Thus, for instance, there is no reason to believe that people should even uniformly be averse to pain. A wounded soldier on a battlefield might well revel in feeling pain both because it reveals to him that he has survived the wound and because pain is a harbinger of discharge from the front. See id. at 102. At a minimum, it would appear necessary to note that people seek to avoid the suffering that usually (but not invariably) accompanies pain rather than pain itself. This distinction between pain and suffering is explored in some detail in Eric J. Cassell, The Nature of Suffering and the Goals of Medicine, 306 NEW ENG. J. MED. 639 (1982).

to those gained from (what some might dub) excitement or "pleasure?" How do we judge the satisfactions that accompany the mild frustration of sorting through something puzzling or new? An increase in comfort may be good, but if it is accompanied by a decrease in "pleasure" it is unclear whether we have witnessed an improvement or decline in a meaningful welfare level. We may (arguably) be able to measure each, but not have any good idea what to ask people to tell us about the "balance."

54. This contrast dominates Scitovsky's well-known social/psychological theory of the growing discontent in American culture. See Tibor Scitovsky, The Joyless Economy (1976). My point here is not to argue that Scitovsky's argument is right, but simply to note that he raises a not-readily-refuted argument: we may readily describe two unambiguously positive hedonic states (in his case comfort and pleasure) whose simultaneous attainment may not be possible, and whose "relative value" is not readily commensurable. Id. at 59-79. One can tell, in other words, that one is comfortable (or stimulated) but there is no apparent psychological mechanism to balance achievement of each end state.

Scitovsky argues in the book that people essentially require two quite distinct forms of "goods"—those that satisfy basic physiological needs that cause unpleasant levels of arousal (needs, e.g. for food, drink, sex) that can indeed be satiated (giving rise to "comfort") and those required to alleviate boredom (creating pleasure), through optimal novelty and stimulation. In his view, feelings of comfort and discomfort (related to the level of arousal) are not commensurate with feelings of pleasure (a function of changes in arousal level). See id. at 61. He argues, moreover, that positive and negative experiences are not on the same scale, to be balanced against one another, but refer to wholly different sorts of experiences, but I want to leave that form of incommensurability aside. Id. Pleasure achieved as a by-product of need reduction is fleeting, though; since it lasts only so long as the change from discomfort to comfort lasts; seeking pleasure through need reduction, then, rather than through stimulation is a no-win hedonic strategy. Id. at 62. Scitovsky argues that people tend to choose comfort (the reduction of unpleasant arousal) over stimulation, in part because advanced economies are so readily able to meet demands for comfort, see id. at 10, and in part because (1) we are instantly rewarded when we satisfy demands for comfort, while the impact of doing so on long-term pleasure is revealed to us only gradually and people tend to respond to instant, rather than complex long-term reinforcement, and (2) the process of satisfying a need is pleasurable and we tend to want to repeat the process of eliminating a want, though as we reduce wants we are less able to do so. See id. at 66, 71-77 ("[T]he satisfaction of a need gives both pleasure and comfort. But the continuous maintenance of comfort would eliminate pleasure, because, with arousal continuously at its optimal level, there can be no change in arousal towards the optimum."). Moreover, the choice is poor in a host of ways meeting demands for comfort is more resource intensive and thus environmentally problematic, more prone to be met by mass production/consumption which depletes the capacity to experience novelty, raises the (relative) price of novel goods and experiences, generates fewer positive consumption externalities than "stimulating" consumption, and
Similarly, psychologists who study the functions of affective states often sharply differentiate positive and negative states that others might treat as unified. A substantial number of emotions researchers would argue that one's mood can be good (or bad) though one is unhappy (or happy) because "mood" (unlike emotions like "happiness") is dominantly a function of the individual's sense that his resources are adequate to meet his current demands while emotions (arguably) represent more momentary responses to hedonically relevant present events. (Thus, in good moods, most goals seem attainable and people are likely to engage in goal-directed behavior; in bad moods they become self-focused and passive). Once more, whether improvements in mood are commensurable with improvements in (more momentary) emotion is hardly obvious.

---


56. To some extent, of course, too, the question of whether people generally in good moods are "better off" than those who have accumulated more positive states depends in significant part on a (weak) theory of human flourishing. Those who believe that mastery and sense of control are key human attributes are arguably more likely to value mood compared to pleasure, which can somewhat more typically be achieved more passively. On the other hand, one can readily imagine a group of observers who devalue good mood because it (arguably) leads to the repressive denial of negative affect, and such observers might (conceivably) believe that self-delusive "happiness" is less valuable than happiness that survives reflection.

I return later to the controversial psychological literature on what is conventionally dubbed "depressive realism"—the tendency for depressives to be less deluded about a variety of circumstances and prospects in their lives (most
In much the same way, we might believe that so-called "pleasures of the mind" are indeed profoundly incommensurate with "pleasures of the body," notwithstanding the legitimate suspicion that such distinctions are just brain-workers' elitist yammering.\(^5\) In one view, people have domain-specific responses of attraction to stimuli. Typically, according to proponents of this view, the main sources of bodily pleasures are our sheath of skin and the entrance and exit holes in our bodies: bodily pleasures may be tonic (time extended) pleasures from specific stimuli (caresses, flowery aromas, sweet foods) or relief pleasures (like sneezing). The relief pleasures involve almost no cognitive or meaning-giving activity while the tonic pleasures are both direct and meaning-mediated (though still distinct from the pleasures classed as "pleasures of the mind" in ways I am about to describe).

Pleasures of the mind, though, are distinct both in terms of their manifestations, uniqueness to humans, duration, level of volition associated with them, and fundamental origin. Thus, they are not inexorably accompanied by a distinct physiological response (including a distinctive universal signal such as a facial expression); they are far less likely to be present in all primates; and they are typically time-extended; their occurrence is more typically deliberately sought out. More profoundly, the "emotional" sequence that precedes these pleasures is typically complex in ways that even tonic body pleasures are not (in that it is not so obviously a sequence of unambiguously positive events). Pleasures of the mind seem to occur when expectations are violated, causing autonomic nervous system arousal (an unpleasant anxiety not only their level of control over random or chance events) and (most particularly negative) judgments others make of them. I will do so in discussing why hedonic psychologists might distrust the widespread claim by "preference utilitarians" that we can trust that fuller information will permit us to reach more truly satisfactory states. For now, though, I just wanted to note the possibility that some might discount the importance of "good moods" because they lead to (or at least seem to be associated with) self-deception. (Assuming the two are correlated, it is plausible that good moods cause us to become unrealistic; it is also possible that one can only be happy if one represses lots of information, especially about our fundamental helplessness and inability to influence critical aspects of our environment).

\(^5\) For a review essay expounding this position, see Michael Kubony, *On the Pleasures of the Mind*, in *WELL-BEING*, supra note 2, at 134.
state), followed by a search for interpretation that culminates in arousal reduction. In one view, the classic pleasure of the mind arises from the sequence of emotions (perhaps a high, but not too-high peak of arousal followed by a satisfactory ending).  

Thus, in this view, what gives rise to intense enjoyment of music is its complexity or unfamiliarity. We first "appreciate" music when we internalize its stylistic structure, develop expectations about what we are likely to hear, and get pleasure when they are realized. A new piece in a somewhat familiar style fits our general expectations but the details are jarring; thus, one begins with expectations that are violated by a complex or novel piece, causing a search for a sensible narrative interpretation of what one hears. The partial match causes satisfaction; the partial mismatch precludes boredom (at least till a piece becomes "too familiar"). Again, my main point for now, is that it is not at all clear how to elicit information about "how happy" subjects are given the possibility that the varieties of "happiness" are so distinct. Questions can emphasize short-lived body pleasures or (distinct? more significant?) pleasures of the mind, and if we try to gather

58. Note that the "additive" moment-based schema that Kahneman employs for measuring objective happiness over a period cannot directly account for the significance of sequencing; it should do so indirectly, though, if a particular sequence of (past) emotions gives rise to a positive emotion (or negative one) that would not have occurred had prior events occurred in a different order. See Kahneman, Experienced Utility, supra note 5, at 678 ("[T]he contribution of an element to the global utility of the sequence is independent of the elements that preceded and followed it. . . . In a moment-based treatment. . . . the elements of the sequence that is to be evaluated are not events but rather moment utilities associated with events. Because all the effects of the order of events are already incorporated into moment utilities, the order of these moment utilities no longer matters."). The (practical) scaling problem is a hugely difficult one though to solve in these cases. Unless we have a good way of accounting for the ultimate (high?) intensity of the final experience, those who experience "pleasures of the mind" will appear unhappier than they are, because some of their instantaneous reports should be negative.

59. For an early, less explicitly psychological argument to this effect, see LEONARD B. MEYER, EXPLAINING MUSIC (1973). For a revised, more complex and explicitly psychological reinterpretation, see EUGENE NARMOUR, THE ANALYSIS AND COGNITION OF MELODIC COMPLEXITY: THE IMPLICATION-REALIZATION MODEL (1992).

60. See William W. Gaver & George Mandler, Play It Again, Sam: On Liking Music, 1 COGNITION & EMOTION 259 (1987).
information about both, the scaling problem may well be insurmountable.

3. Integrated Versus Dissolved Views of Pleasure and Pain: Reflecting on the Pain Duration Experiments and on Hedonic Adaptation. The problematic nature of the basic claim in the “new hedonics” literature is perhaps most clearly revealed if we analyze what strikes me as the strongest data supporting the proposition that the total amount of “pain” and “pleasure” each of us has felt is a real fact that can be misreported by a good faith subject: the pain duration experiments. Recall that the experiments demonstrate that increasing the duration of pain while lowering the end point level of pain leads to a decline in the reported level of pain for the whole incident. It seems to violate certain intuitive logical/mathematical principles (usually referred to as violations of “dominance” rules) for a

61. Recall that those who view this as a simple reporting error believe that the root of the reporting error is the fallibility of memory. Unable to recall and sum all of the moments of pain in a long sequence, people simply recall what is easiest to recall—the last moment and the worst moment—and assume that when they report on those moments, they are probably reporting the entire experience accurately. In this sense, the “error” peak/end reporters make is much like the error that subjects make generally when they make global evaluations of a set by reference to salient prototypes, ignoring base rate information. This interpretation is the one proffered in Charles A. Schreiber & Daniel Kahneman, Determinants of the Remembered Utility of Aversive Sounds, 129 J. EXPERIMENTAL PSYCHOL.: GEN. 27, 28-30 (2000), drawing on Daniel Kahneman & Amos Tversky, On the Psychology of Prediction, 80 PSYCHOL. REV. 237 (1973), on base rate neglect more generally.

It is possible, of course, that peak/end reporting is an “error” but that it is nonetheless functional to attend both prospectively and in real time (not just in recalling and reporting) only to the worst and final pains one will suffer during an incident. The reason that we recall peak/end data may be that these moments are especially rich in self-relevant information. Peaks tell us something relative to our capacity to withstand pain (so that a relatively low peak means that the episode presented—and such episodes will present—few coping problems) and endings permit us to know where we stand with certainty. “Peaks and ends... earn their privileged status because they carry more personal meaning than other moments.” This account of peak/end reporting is taken from Barbara L. Frederickson, Extracting Meaning from Past Affective Experiences: The Importance of Peaks, Ends, and Specific Emotions, 14 COGNITION & EMOTION 577, 588 (2000). I suspect, though, that Frederickson’s view, like the one I am about to explore in the text, is that peak/end reporting is not an “error” at all, but that “reflective” judgments of chunks of time encoded as discrete “incidents” are (at least) just as “real” as moment-by-moment additive judgments.
person to feel he was happier if he added a poorer experience to an identical experience. But that intuition depends on believing that happiness is simply the sum of aversive and pleasant physiological momentary experiences. Just as it is not clear that happiness is the reflected stable understanding and evaluation of an experience that persists across time, or, as skeptical economists/preference utilitarians would declare, that utility is simply the instantiated outcome that we reach if we obtain an end-state that is chosen under conditions of perfect information, so this assertion is hardly uncontroversial.

While I will largely focus my attention on the peak/end problem, the arguments I will raise in discussing whether peak/end pain reporters are “in error” could be raised in other psychologically significant contexts as well. For instance, if one carefully explores the relationship between “certainty” and hedonic adaptation, I think one will also find that subjects typically violate the simple additive or “mathematical” view of hedonic satisfaction that the early Kahneman posited as the sole rationally defensible account of hedonic satisfaction.

The basic, robust finding of the uncertainty/adaptation literature is that people may adapt better (and hence “feel” better, at least in some meaningful sense) when they are sure something bad has happened than when they are largely (say, 95%) sure but still see a way out. In this regard, those who receive bad HIV reports or a firm diagnosis that they have Huntington’s disease seem “better off” than those not yet informed, but highly suspicious, of their status. Under one view, of course, the person with a 100% chance of something bad happening is just like the person with a 95% chance except that as to the last 5%, he is in worse shape. One way of putting that is that state X is being 95% certain that one is HIV positive, and that state is present both in those who are certain that they are HIV

---

62. See generally Jeffrey M. Moulton et al., Results of a One Year Longitudinal Study of HIV Notification from the San Francisco General Hospital Cohort, 4 J. ACQUIRED IMMUNE DEFICIENCY SYNDROMES 787 (1991); Jason Brandt et al., Presymptomatic Diagnosis of Delayed-Onset Disease with Linked DNA Markers: The Experience in Huntington’s Disease, 261 J. AM. MED. ASSN 3108 (1989) (although sample sizes are small, people seem to cope well when they receive alarming genetic test results, particularly if they are anticipated to some extent).
positive and those who are not yet informed but are fairly sure that they are HIV positive. If state \( Q \) is “a lottery ticket with a 5% chance that I learn information about my HIV status and the information that I learn is that I am HIV positive” and state \( R \) is “a lottery ticket with a 5% chance that I learn about my HIV status and the information that I learn is that I am HIV negative,” one plainly prefers \( R \) to \( Q \), but one may not in fact prefer the sum of \( X \) and \( R \) to \( X \) and \( Q \). The reason, presumably, that the mathematically rational dominance model seems to fail is that it does not seem to deal with the “integrating” experience agent—the “ego” that adapts to bad news and somehow goes on, or suffers from prospective anxiety so long as he possesses lottery tickets rather than certain information.

In a sense, one could argue that the presence of such an “integrating” agent makes a whole slew of end-states “complementary” that would not otherwise appear to have such a quality. Obviously, perfectly conventional rational choice theorists recognize that a chooser \( C \) could prefer \( X + Z \) to \( Y + Z \) even if he preferred \( Y \) to \( X \) if \( X \) and \( Z \) were in some sense complements—e.g., if \( X \) were a condiment that went especially well with some food \( Z \) though it is not very tasty on its own—but we typically think of complementarity

---

63. Alternatively, one could frame this point by noting that it is clear that the subject prefers state \( E \), HIV- status, to \( D \), HIV+ status: but that ordinarily implies that the expected value of a 95% chance of \( D \) plus a 5% chance of \( E \) is valued more than a 100% chance of \( D \). The fact that this set of preferences violates ordinary rationality conventions can be seen if we imagine the state \( E \) being “winning” $1 million—parallel to HIV negative status—and state \( D \) being winning only $100. It would plainly be better to have a 95% chance of \( D \) and a 5% chance of \( E \) than a 100% chance of \( D \), and we would think we would be able to infer that from the fact that \( E \) is preferred to \( D \).

The conventional rationality principle that this violates is Savage’s “sure thing” principle—the principle that if one is offered a lottery \( X \) and lottery \( Y \) which differ only with respect to the fact that \( X \) contains prize \( A \) as one prize and \( Y \) contains \( B \), that one rationally “must” prefer \( X \) to \( Y \) if one prefers \( A \) to \( B \). My main point for now is the fact that Savage’s sure thing principle is not true in relationship to adapting to very bad news suggests that rational addition/dominance models may fail more generally, not just in the pain duration cases. For the initial formulation of the “sure thing” principle see, L.J. Savage, The Foundation of Statistics 21-26, 76-82, 103 (1954). For an excellent discussion of the principle and a critique, expressing strong doubts that human beings invariably make choices consistent with the “principle,” see Daniel Ellsberg, Risk, Ambiguity and the Savage Axioms, 75 Q.J. Econ. 643 (1961).
in consumption as a special, minor case, that we would expect to find when the goods were *used* together, in a physical sense, rather than *reconstructed*.

Now, I am not claiming that the "integrating person" here—the "one" who can engage in adaptation or suffer anxiety—is precisely like the integrator of pain information—the one who reflects on how painful he found some "whole incident." In the HIV-example, both the adapting party and the still-uncertain party are each plainly having new, ongoing hedonically-charged experiences (e.g., anxiety, resignation and adjustment) rather than simply attempting to sum feelings that occurred in the past. Thus, Kahneman (would plainly) argue (quite reasonably) that the additive, moment-based model is adequate so long as we recognize that the uncertain party has a hedonic reaction not just to the information set available to him but to the *presence* of anxiety and the informed party has a hedonic reaction not just to the information set but to his *reduction* of anxiety.

Kahneman implies, in treating the peak/end reporter as "mistaken," that happiness judgments must be conceptually disintegrated. Every judgment of happiness or suffering is the judgment of happiness or suffering at a particular moment. And what ecological momentary assessment does is ask respondents, in essence, how they are feeling now. He naturally acknowledges that suffering or happiness experienced at a moment can be *caused* by past or future events. We can feel currently unpleasant dread (or feel good) in anticipating the future, we can joyfully reminisce or relive horrors. But what we cannot do, in his view, is "experience" anything but present-tense reactions. If Patient X feels bad now because he *recalls* a painful procedure, that is real; but if X simply says he felt bad

---

64. I believe he might now qualify the statement in the text by noting that while "happiness" judgments should be disintegrated in the way I imply, there is more to "well-being" than happiness and (perhaps) some aspects of well-being could, at least in theory, be discerned only by an integrating agent. Once more, I discuss the "purer," earlier view both because it is intrinsically important and because I am not absolutely certain whether the later Kahneman really believes that aspects of the "fuller" vision of well-being require us to look to reactions that cannot be properly understood as momentary reactions (even if they are not momentary approach/avoid reactions).

65. Again, to emphasize the parallel with adaptation to bad outcomes that have become certain, he may feel good now though he has heard bad news
during the procedure, that judgment is entitled to no independent weight. It is merely a (correct or incorrect) observation or reconstruction of an objective state.  Note, in this sense, that X's judgment of the hedonic quality of the experience is just like a third party's judgment of X's experience. The fact that he is interpreting his own life is of no moment.

It is worth reflecting on the family resemblance of this view to Parfit's argument on personal identity. As a prelude to his defense of utilitarianism against the commonplace accusation that it does not take adequate account of the separateness of persons, Parfit goes to elaborate lengths to show that choices made to benefit one's own future self are scarcely distinct from choices made to benefit others. To do because his anxiety is reduced or he is now meeting his (lowered) hedonic expectations.

66. If the person who has seemingly adapted to the “bad news” state misreports happiness not because he has become happier but because he reports the same lower level of happiness as a better state once his expectations of happiness decline, then, in Kahneman's terms, he has not really adapted hedonically. Kahneman, in fact, hypothesized that people do not shift the degree to which they react positively or negatively to experiences as a result of having had good (bad) experiences, but simply report the same range of good/bad reactions as better or worse depending on the degree to which they expect high or low levels of good (or bad) reactions. See Kahneman, Objective Happiness, in WELL-BEING, FOUNDATIONS, supra note 5, at 13-14; Kahneman, Experienced Utility, supra note 5, at 685-89.

67. See DEREK PARFIT, REASONS AND PERSONS 199-347 (1984). Even if Parfit convinces us that persons are less integrated than we typically believe, this may do far less than he hopes to revive the argument that utilitarian ethicists legitimately ignore the separateness of persons and focus on maximizing the occurrence of hedonic pleasure, regardless of the locus of that pleasure. For one of many sustained arguments that Parfit's argument is at core a non sequitur (as even a presumptive defense of utilitarianism) see Bart Schulz, Persons, Selves, and Utilitarianism, 96 ETHICS 721, 732 (1986) (stating that Parfit can claim no more than that his Reductionist view of persons makes it less likely that we will attend to the distinction between individuals, not that it will obliterate it).

Moreover, Rawls argues that so long as stable societies can exist that govern themselves using a more Kantian conception of persons—in which people view themselves as autonomous and take responsibility for their fundamental aims over a far greater period of time than anything we could call an “experience” would last—the fact that a different view of persons might also be metaphysically plausible is of little moment. Rawls states:

One can imagine people who are hedonistic and individualistic; their lives lack the connectedness and sense of longer purpose needed for a Kantian view to work. But that this may happen under certain conditions so far shows nothing about what is desirable from a moral
this, he must dissolve conventional ideas of identity, arguing, for instance, that we should be unconcerned about our own deaths if (under certain science fiction assumptions) another being with our memories and intentions still existed. What links fundamentally separate selves over time is not some “further fact”—like some extra-bodily Cartesian ego to which all of the events in “a person’s” life happen—but more contingent psychological links (memory of the past, intentions towards the future, similarity in beliefs).

Point of view. There is no degree of connectedness that is natural or fixed; the actual continuities and sense of purpose in people’s lives is relative to the socially achieved moral conception. Thus the essential point is whether the well-ordered society corresponding to a moral conception generates in its members the necessary continuities and sense of purpose to maintain itself. . . . [A] utilitarian view would be supported by the general possibility of discontinuities only if social theory showed that in the case of other conceptions the requisite connectedness could never be brought about.


68. Again, I am not sure how much of this philosophy of mind material is my cup of tea (or was at the moment I wrote this, when I was a different person)? But in the fully dissolved view, virtually all of our moralistic social practices disappear. It is bizarre, for instance, to punish someone for something he already did (or, to focus on the effort to dissolve the distinction between the self and others that Parfit is pushing in his defense of conventional utilitarianism, no more rational to punish a criminal than his son for his misdeeds). This would appear to be the case because “he” is no longer the criminal. Nor would it be especially plausible that deterrence could work because no one has any good reason to plan, to avoid future consequences that (after all) really occur to someone else. Parfit notes that as memory fades, we are less connected to our past selves—and in his view, less justly punished for that past “self’s” conduct, just as we would be less morally bound to keep “his” promises. He is relatively ambivalent about the effect of thinner views of identity on punishment and promise-keeping in PARFIT, supra note 67, 323-39, but by the time he had written, Derek Parfit, Comments, 96 ETHICS 837 (1986), his views on the issue were far firmer, and more cleanly reject the conventional moralistic positions.

Just as some are dubious that the philosophy of mind has much to do with the battle between utilitarianism and Kantian theories, it is also questionable whether conventional “moral practices”—e.g., of promise-keeping or punishment—really depend on the metaphysical presuppositions. The fact that we might have trouble justifying punishment in a case in which our commonsensical judgments about the continuity of a person—in body terms, memory terms, “further” terms—were most strained (e.g., by memory loss, or science fictional transportation of memory to another body) tells us little about whether the metaphysical identity assumptions actually ground the routine practices. For an elaborate argument to that effect, see Mark Johnston, Reasons
and Reductionism, 101 PHIL. REV. 589 (1992). See also Samuel Scheffler, Ethics, Personal Identity, and Ideals of the Person, 12 CAN. J. PHIL. 229 (1982). Parfit indeed came to believe both that punishment for past misdeeds was inapt and that it was senseless to attempt to compensate someone for causing him pain, since one would inevitably "compensate" a significantly distinct being. But the point strikes me as fundamentally somewhat obtuse. It is hardly clear that one could "deserve" punishment—under any conventional account of desert, even if one were punished moments after one took bad acts—unless we conceived of "evil choice" as something performed by integrated individuals, capable of devising meaningful life plans. The dissolved "later" person is not so much non-punishable—as Parfit suggests—because he is not the blameworthy one, (though such a person did exist, albeit fleetingly?) as he is non-punishable because no one could justly be punished in a world in which lives were dissolved into (significantly) disconnected moments, with no transcendent agent able to choose a life plan or moral posture. PARFIT, supra note 67, 323-26.

Similarly, the point about compensation appears to me a non sequitur. Assume that I agree to allow you to harm me in some fashion in exchange for a later payment. What justifies your causing me harm may well be my agreement that you can harm me (my belief that I am not so dissolved that I find it preposterous to sacrifice anything for the sake of some disconnected future me, coupled with my "right" to waive my prima-facie, but alienable, rights against certain forms of harm-causing). See Christine Korsgaard, Personal Identity and the Unity of Agency: A Kantian Response to Parfit, 18 PHIL. & PUB. AFF. 101, 130 (1989). Parfit cannot sensibly argue that such concern for the distant future is so irrational as to vitiate the ostensible consent, if he is to allow exchange at all, for in any complex exchange action, the parties will inevitably reap benefits at a particular point in time. Thus, assume the actor agreed to some smaller level of instant gratification/compensation in the more immediate future. At some level, any delay poses the same conceptual problem as long delay (why should the only sentient being—the present tense I—care about another being, some future sentient being who remembers lots about me), but at another level, he could justifiably kick himself for having overvalued a "self" that will soon be significantly gone (merely "remembered"). If there is no reason for the "present" I to care about the future, there is no reason for him to care about the present either, knowing it will soon be past. I take this to be one view of Korsgaard's point that:

It is misleading to ask whether my present self has a reason to be concerned with my future selves. This way of talking presupposes that the present self is necessarily interested in the quality of present experiences, and needs a further reason to care for more than that. But insofar as I constitute myself as an agent living a particular life, I will not in this way oppose my present self to future ones. Id. at 126-27. In arguing for compensation in situations in which prior agreement was not obtained, it seems to make more sense to think about mimicking the agreements that might have occurred than to assume that no such agreements ever would have.

More pointedly, I think, it is hard to understand such future-regarding practices as disciplining your children for the sake of their adult selves if the "child" we love is as thoroughly disconnected from her adult embodiment as Parfit suggests. (If he is right that persons are linked solely by memory and intention, then children and adults are scarcely linked, especially if one is
Now, of course, Kahneman and Parfit’s views hardly jibe. Parfit is arguing that the question of whether Person P is happy (over time) is scarcely meaningful. We can (and should) sum the experiences of all persons as readily as we can sum the happiness of P. Parfit aspires to make us more impersonal, to get us to focus less on persons and more on experiences. Kahneman, on the other hand, wants to maintain the notion that (some conventionally defined) P’s happiness (summed over time) is significant enough to try to measure, but it is never clear why we should wish to do that unless that person P has some transcendent identity. 69

talking about very young children. Adults will typically not remember all that much of childhood nor does the child have many firm intentions about “his” adult future). A variant of this point is emphasized in Susan Wolf, Self-Interest and Interest in Selves, 96 ETHICS 704 (1986).

69. It is interesting in this regard to note the relationship between two questions that are often radically separated in the philosophical literature on identity, the literature on what makes a person a person. One question we might want to answer is what distinguishes persons from non-persons (this generally includes a discussion of why we might be more concerned in a moral sense with those we think of as persons). Another issue we are interested in is what makes P the same person at time 2 as he was at time 1. It might be helpful to note that the ways in which we are the same “person” across time might be the source of what is (morally) special about us. We are meaning-giving agents over time. In this view, what makes (most of us) care more about the pain and pleasure experiences that occur to people than to animals is that people have conceptions of more and less pleasurable lives, as well as the capacity to sustain relationships and plans over time. As Wolf puts it:

This provides a reply to a proposal sometimes suggested by Parfit that we should care more about the quality of experiences and less about the persons who have them. For the value of persons is not, as this proposal would suggest, dependent on their ability to have such momentary experiences. . . . If the reason we care about persons is that persons are able to live interesting, admirable, and rewarding lives, we may answer that time slices of persons, much less experiences of time slices, are incapable of living lives at all.

Wolf, supra note 68, at 709.

There is also, of course, a view—from my obviously limited understanding, I take this to be the commonplace view among evolutionary biologists—that persons are typically bound across time for a reason that makes no reference at all to what is or might be unique about people (compared to other animals, plants or even microbes). The human body (like any living organism) is just a “survival machine” containing those genes (whether acting alone or in quite complex combination) that have become most prevalent over time because they reproduce themselves most frequently. The self in that view is integrated over time precisely to the extent that behavioral integration increases the chances of gene replication. (Thus, genes that produced disintegrated behavior of the extreme form: “Prefer pleasure in pre-reproductive phases of life that posed enormous risks of dying before reproduction because the selves that would be
Without such an identity, balancing pains and pleasures over a lifetime is not a sensible strategy. Parfit seems right that self-centered indifference to one’s future—in the sense that it is of no more importance than anyone else’s life—would be appropriate if it were not in some sense yours.

In a sense, then, Kahneman tries to adopt a *middle position* in between the conventional position on identity—what gives us identity is our capacity to integrate experience and it is the integrator of experience that we care about both as a moral entity and as the subject of welfare measurements—and Parfit’s radically dissolved position—that there is no such integrator and therefore we should care less about the locus of experiences (as deontological rights-based theorists do) and more about the experiences themselves. I believe that Parfit’s view of the dissolved self—which Kahneman’s is actually foundationally closer to—is very difficult to defend. 70 What I am alive to reproduce are not really ‘me’” would probably not have survived). By and large, one would assume that selves would be pretty integrated as carriers of the replicating genes—and that the survival of this integrated self would have priority over the survival of others precisely because non-related others would typically not be vessels for the replicating genes. At the same time, though, selves are not “profoundly” integrated. For instance, a replicating gene would have been just as likely to become prevalent if it produced behavior that manifested less care for its future self than for the survival of a high number of gene-sharing siblings. Moreover, while genes replicate successfully enough to become prevalent in part because they dictate behavior in the organisms in which they are embodied—a view that implies that the “continuous body” is *meaningful* as the repository of genes and the vehicle for their survival—genes may survive because they impact the environment outside the body as well (e.g. by secreting chemicals that alter the behavior of would-be predators), so that the idea that the body containing genes has some “unique” genetic significance is surely overstated. The usual accessible account for us lay readers trying to make sense of these views is RICHARD DAWSON, THE SELFISH GENE (2d ed. 1989).

70. I make a brief argument in the text to that effect, but also wanted to explore another way of getting at why Parfit’s views on this matter are so controversial. I think it could be helpful as well in judging both the plausibility of the view that well-lived lives are “integrated” rather than dissolved and the analytical distinction between Kahneman’s “moment addition” conception of welfare (associated with “thin” views of persons) and “reflective” views (associated with “thicker” views of persons) to reflect more on some of what strike me as the most powerful critiques of Parfit’s work. In this regard, it would be especially helpful to interrogate the meaning of Parfit’s claim that what links us with our past “selves” is nothing more than (weaker or stronger) memory—just as what links us to the future is weaker or stronger intentions. The notion that all that links us with our past is memory is subject to the (familiar) critique—apparently dating back to Berkeley—that the claim is circular. If X and Y each “recall” having seen the Eiffel Tower from a passing
Metro train, what could make X's belief a memory and Y's a delusion? X's is a memory if and only if the event actually occurred to him; but to know whether the event occurred to him, we must have some separate account of his identity (beyond that he is an entity linked by memories). Thus, the effort to define identity in terms of memory (Y is the same person as X only insofar as he recalls events in X's life...) fails, for we would then need an account of identity to define memory. For a far fuller discussion of this point, see Marya Schectman, Personhood and Personal Identity, 87 J. PHIL. 71 (1990).

For my purposes, Parfit's effort in REASONS AND PERSONS 219-226, to avoid the circularity problem by invoking the idea that people can have what are generally called quasi-memories (pictures of events that actually were caused by a person's experiences rather than imagination) without invoking the idea that it was their experience that caused the sensory picture) are largely irrelevant. What I need to focus on is what strikes me as his fundamentally narrow understanding of memory, an understanding that permits him to maintain his very thin view of persons, and to ignore the ways in which "identity" might be nothing more than the features of personhood that make it thick rather than thin.

My argument about the inadequacy of Parfit's account of memory does not depend directly on the circularity critique nor, therefore, is it answered by the invocation of quasi-memories. What is far more significant in my view is that a reasonable phenomenological account of memory (and intention) entails thick, not thin, identity, the capacity to integrate events over time and to narrate both a past and a set of life plans that transcend highly particularized sensory memories (or detailed micro-plans).

Most actual recollections are, first, not fully detailed single sense impressions, but meaning-infused summary impressions of events, and even precisely recalled sense impressions may make sense only given a backdrop of far vaguer personal recollections (which, quite critically, cannot themselves be reduced to direct sense impressions). The most suggestive way of putting it is that no one but me could have my memories; the "memories" are a product of my construal/reduction/interpretation of a host of life experiences whose sense-impression details are largely lost. Thus, one of many accounts of the "thicker" version of me is "that unique person capable of having my memories." If we try to follow Parfit's science fiction experiments and imagine "memories" taken from X's head and transplanted into Y's, it seems that Y will either find the traces of sensory memory confusing and incomplete (that is, he will not really have more than confusing bits of X's memory), or if he gains enough (of X's meaning-giving) context so that the memory is not vague and confusing, incredibly unsettling because it will appear (from the inside) as he thinks that he has lived an unfamiliar life or is having a delusion.

Assume, for instance, that you receive (by Parfit's sort of sci-fi memory-transplant) what strikes me as an atypically vivid memory from my (disturbingly) long-past adolescence, a "memory" of walking in a park near my parents' home with a girl I had a long-entangled, remarkably dysfunctional relationship with. My "memory" of the physical appearance of the park is almost surely suffused with my separate knowledge of the park (which I have surely walked through more than a thousand other times). Thus, if all you received were a sense impression not reinforced by separate knowledge, you would receive far less vivid detail than I actually can summon (though it might conceivably be in some limited senses a more 'accurate' portrayal of the park in
1969, since what I now experience as my 'memory' may be distorted by subsequent changes in the park). If the park seemed familiar to you (is part of what Parfit understands as transplanted "memory" the appropriate sense of familiarity or is it 'merely' the sensory recall?), though, that familiarity would doubtless be unsettling to you rather than nostalgic (if you have never been there, the ability to recall a physical space you walked through briefly more than thirty years earlier would itself be a source of disquiet, not context). Similarly, my recollection of the girl's appearance in the memory is doubtless colored by innumerable interactions (both before and since the event), and is sensible, in many ways, not as a sensory memory but as an interpretation of how I had come to believe she looked at that point in time given my feelings for her. Here is another version of the same problem with the sensory impression view of memory: when I recall the girl, she in some sense "appears" to be (more-or-less) "my" age. (This is true only in some very complex, interpretive sense. She is "my" age in part because I do not encode myself as having any particular age when I envision her age in relationship to my own. It is clearer to me, then, that she does not look "older" or "younger" than that she looks any particular age at all. Thus, older and younger are not interpreted, in my mind, as older or younger than any particular pictured, embodied age-type. I encode her as neither older nor younger dominantly in the social meaning sense: e.g., she does not seem "different" from me or inappropriate). If, though, I received a wholly sensory image of the transplanted memory of some girl who looked just that same age from some Parfit-ian science fictional memory-donor, she would surely look like a kid to me, look essentially like someone my teenage sons go to school with. Encoding her physical image in that way, though, would distort my meaning-infused memory of the interaction.

The detailed sense of my (highly) anxious and miserable (but also partly ecstatic) state would be senseless without thoroughly "understanding" the complex history of the relationship, most of which I could not encapsulate in any other set of particularized, detailed sensory memories that I could readily "ship out" for transplant. (I could tell you, in some general sense, that we had innumerable phone conversations of a certain sort, but I cannot recall any conversations in detail, let alone verbatim, as direct sense memories. Certainly, I could use—inevitably inadequate—words to describe the nature of these phone calls; to the degree you "understood" and integrated my account, you would do so both imperfectly and in a fashion that probably reflected your own experiences more than mine). I am not linked by (more or less accurate) memory to the person present at that event. Memory of the event is something that only (a thick, meaning-giving) "I" can provide. That thick "I" need not be a mystical transcendent Cartesian Ego; but it may be the vector resultant of such complex sensory and interpretive systems that it cannot be re-dissolved into any set of simple components, with certain sorts of knowledge of one another. I am fairly confident that there is at least a loose family resemblance between this argument and a series of far more complex arguments made against the possibility of artificial intelligence by people like Dreyfus, but I totally disclaim the ability to explore that angle at all. See HUBERT L. DREYFUS, WHAT COMPUTERS CAN'T DO: A CRITIQUE OF ARTIFICIAL REASON 197-217 (1972).

Another way of putting this point is that Parfit gets very little by saying that persons are dissolved strings of experiences connected by "mere" memory because there is not much about "memory" that is "mere;" memory itself is one of the main manifestations of thick-meaning-giving, non-dissolved experience.
emphasizing though is that Kahneman's middle view is the least defensible. Arguments for caring about an agent's welfare depend on there being an integrated and integrating agent to care about. In this regard, it is important to emphasize what I find the strongest basis of the standard attack on utilitarianism that Parfit was trying to deflect. The strongest basis of the attack on using the utility maximization principle without regard to persons need not be based on any particular theory of the rights of particular individuals against particular other individuals, it is at core grounded in the fear that utilitarians adopt the position they do because they have an overly thin account of persons, that they view them simply as "containers of valuable experiences." If societies are best construed as nothing more than a series of such containers (unworthy of personal identification), then it indeed makes sense to construe each "container" as a container for mini-containers. If persons are unworthy of personal identification, why attend to the claims of the aspect of the Parfit tells the story as though I would (essentially) have your identity if I read a log of the life events you believe most critical to you; but the log account of memory is in my view rather transparently unpersuasive.

71. Thus, in my view, utilitarians are not necessarily vulnerable to the objection that they pay inadequate heed to the separateness of persons if they advocate, say, punishing an innocent person to prevent some unjustified harm to other persons. Such a moral scheme could be grounded in a (relatively reduced) sense of the importance of not directly harming the innocent (that is to say, a lower level of agent-relativity), or a (relatively increased) sense of our duties to protect others against indirect harms. The usual notion is that the anti-utilitarian intuitions on these issues depend on what is referred to as "agent relative" moralities (moral postures in which the integrity and moral status of the actor, as well as the acts, has some significance). My own intuition is that it is easier to see that agent relative moralities are inevitable—and that the fact that we might be inevitably agent-relative is an important part of our capacity to act morally—than it is to defend the abstract desirability of agent-relative postures. But in each case we are not (necessarily) merely treating each person as nothing more than a vessel for utils; we might, for instance, take absolutely no account of the differential capacity each party had to experience pain or pleasure in making our decision.

72. For a strong defense of this attack on the utilitarian conception of people, see Rawls, supra note 67, at 5 (1975). See also Amartya Sen & Bernard Williams, Introduction, in UTILITARIANISM AND BEYOND, supra note 45, at 1, 4 ("Essentially, utilitarianism sees persons as locations of their respective utilities—as the sites at which such activities as desiring and having pleasure and pain take place. . . . Persons do not count as individuals any more than individual petrol tanks do in the analysis of the national consumption of petroleum.").
person (his integrative capacities) that most demand such personal identification rather than those most consistent with the idea that each experience has a hedonic valence, to be toted up?

Contrary to Kahneman’s assumption (echoing Parfit), it is not self-evident that there is no significant observing/planning ego, the “subject” of agency, if not some Cartesian ego—who balances pleasures and pains across different periods of life, who evaluates these balances. It is not unambiguous that the “person” who thinks, “I had a less painful experience this past week than I had when I underwent the shorter procedure” is not the most relevant subject; in fact, I find the claim that Kahneman has constructed a fictional locus of utility-summing, never embodied in any actual person, just as facially plausible as the claim that the “integrator” of experience simply misremembers his “genuine” experience.

I do not want to discount the possibility that she has not “construed” the pain differently, that she is simply misremembering pain that she means to interpret as the simple sum of momentary sensory experience, but is cognitively unable to do so; I do not mean to argue that it is transparently incorrect to view Peak-End reporting and the associated duration neglect as an error. It also is possible (though no more than mildly plausible to me) that the integrator who thinks the longer procedure was “better” does so because she feels more or less pain now, in recalling the procedures.

But one cannot discount another possibility: She might simply be engaged in meaning-giving activity, and what she cares about is her construal of her life, her ascription of meaning to events, even purely sensual events. Recall in this regard Frederickson’s intriguing suggestion that reports of incidents emphasize peaks and ends because

73. I also need to note that I am somewhat suspicious of drawing the conclusion that people entirely neglect duration in reporting on the hedonic quality of painful events—even if duration seems to drop out in experimental manipulations in which there are relatively mild alterations in duration. Almost everyone I have ever spoken to who reports on the misery of stomach viruses is quick to note that they are of shorter duration than less painful, but longer-lasting, respiratory flu. Even within the “stomach virus” category, I have noted that people report whether they had a four or twenty-four virus. I recall very few, if any, women (particularly those who chose “natural childbirth”) who have not thought the length of labor salient.
what is most meaningful in painful events is whether they could always be withstood (the peak) and whether they ended as badly as one might have feared while in the middle of the incident (the end). 74 I do not mean to either embrace or disclaim any particular account of why "interpretivists" might summarize "painful periods" as they do: my (more limited) point is that "meaning givers" may not simply summarize by addition of the valence of all moments.

While full-blown duration neglect (if it really existed) would surely seem irrational to me, it would seem (nearly equally) odd for a person to construe her hedonic state without regard to the interpretively-summarized meaning of "conceptually bounded incidents," to believe instead that her welfare could be understood solely by summing the moment-by-moment pain scales. Thus, for instance, it strikes me that nearly all of us would (correctly and rationally) think we had had a worse day if we had suffered "8" (out of 10) pain in two utterly different contexts for an hour each than if we had suffered "8" level pain in one context for two hours, though this implies a form of duration neglect. 75 Now, of course, it is possible to say that having two distinct pain sources in the day should cause more present-tense suffering simply because physical pain is (ordinarily) a signal of prospective danger and the subject should react negatively not only to the sensory pain but the (sensory) anxiety. 76 Even if the source of the pains is not (two distinctly worrisome) anxiety-causing diseases but clearly close-ended events, though, we might well believe that if we suffer two contextually separate traumas in a day, we should experience worry that our self-protective

74. See Frederickson, supra note 61, at 588.
75. I am far less confident that it would be worse to suffer two distinct one hour painful periods six months apart than one two-hour pain bout; it is the concentration of bad events in a short time period, I suspect, that leads people to construe their lives as running amok, either to worry (in some narrow sense) that they face crises in multiple domains or to construe themselves (more broadly) as relatively powerless to ward off bad events.
76. Thus, one obvious explanation for why someone would rather suffer two hours from an arthritic knee than one hour from the knee and one hour from an equally painful toothache is that he might have already known that his knee was in cruddy shape but had no reason to worry that his teeth were failing too. In this (limited) view, the two-incidents pain is worse only because there is as additional pain—dubbed "anxiety" in the text—that is higher in the second case than the first.
faculties are compromised. But at that point, the distinction between “construing” how painful one’s life is and “reporting” the sum of the pains gets very blurry: the source of the additional experienced sensory pain when “things are going bad” is interpretive, not immediately sensory.\(^\text{77}\)

What is vital to recognize in thinking about the possibility of straightforward welfarism, though, is that my preferred view (or any view that privileges meaning-giving) reflects (one weak) perfectionist conception: it treats the capacity to integrate and make sense of disconnected momentary experience as a (the?) key human virtue and treats (at least certain) construals by the person whose experiences are being construed as more authentic than any summing scheme an external observer might impose. At the same time, it disclaims (what I take to be the early Kahneman’s implicit, alternative) weak perfectionist view that people should understand themselves as repositories of good and bad instantaneous reactions, and correlative seek to live the life such a repository would most typically aspire to,\(^\text{78}\) one in which the good in momentary experiences outweighed the bad to the greatest extent possible.\(^\text{79}\)

---

77. Formally, in his more recent work, Kahneman acknowledges that one can construct an objective happiness measure over time only if each moment’s pleasure or pain has the same meaning as each other moments. See Kahneman, \textit{Experienced Utility}, supra note 5, at 679 (“All moments are weighted alike in total utility.”). He acknowledges the possibility then that two distinct moments of equal pain might not count equally, \textit{id.} at 692, but seems to feel that that will occur only in relationship to some (undefined) moments that have meaning outside the pleasure/pain domain.

78. Once more, it is worth noting in that in the more recent work, Kahneman argues that he is simply trying to give voice to the experiencing subject, not to override the world-view of the meaning-given subject whose memories of life are (in his view) conventionally (over)-valued. “The goal of this discussion is not to reject the memory-based view, which is indeed irresistibly appealing, but to point out that intuition is strongly biased against a moment-based view.” \textit{Id.} “Maximizing the time spent on the right side of the affect grid is not the most significant value in life, and adopting this criterion as a guide to life may be morally wrong and perhaps self-defeating as well.” \textit{Id.} at 691.

79. There are some other problems in the “new hedonics” literature that ultimately seem less clearly germane to the question of whether it can hope to avoid “weak perfectionism” but help us better understand that hedonic data is not simply technically difficult to gather, but that the concept of “happiness” and “reported happiness” are very difficult to disentangle.

While it is relatively transparent that ecological momentary assessment will not generate cardinal, fully commensurable and summable utility data, it may be more interesting, I think, if it cannot even generate stable judgments of whether events are positive or negative. Kahneman asserts that reactions to
states are binary, and have a natural zero point: people either like their situation (and wish it would persist), dislike it (and wish it would end), or are in equipoise. But there is neither theoretical nor empirical reason to believe that the state of indifference is a point, as he must assume, rather than a broader or narrower band.

Kahneman does argue that there is a strong theoretical reason to believe that the judgment of good and bad has a single zero point. He argues that judgments of “happiness” are like judgments that a hue is neither dominantly green nor dominantly red, but “white.” Such judgments have a single zero point, even though our ability to discern where that point is located alters with context (those exposed to strong red light just before they are asked to label a stimulus will call light with a higher proportion of red “white”). They are unlike judgments that a line is “neither short nor long” (though, once more, judgments about whether a particular line is short or long depends on the length of lines one has been exposed to prior to the stimulus). His claim is that the length judgments are driven by the need to communicate in relativistic language, while the color (and pain/pleasure) judgments are at core driven by shifts in sensory mechanisms. One way of expressing this insight is to note that respondents to the color experiments think that all the “white” lights they see look the same, while those who describe two lines they describe as “neither short nor long” can still readily differentiate the lines’ length. To put it another way, in Kahneman’s view, pain/pleasure, like color, is bipolar rather than unipolar (length is continuous from shorter to longer, not on one side or the other of a zero point). He also argues that distinct physiological mechanisms (triggering approach and avoidance) are present when we perceive good and bad results (and that we can observe approach or avoidance without having to have subjects communicate their feelings at all). On the approach/avoidance point see also John T. Cacioppo & Gary G. Berntson, Relationships Between Attitudes and Evaluative Space: A Critical Review, With Emphasis on the Separability of Positive and Negative Substrates, 115 PSYCHOL. BULL. 401 (1994).

Kahneman thus put great stock in the idea that those who historically reported surprisingly high happiness levels in the “old hedonics literature” (like accident victims, or the elderly) did not actually “adapt” to their poor circumstances but simply reported happiness relative to diminished expectations. (Kahneman refers to this as a “satisfaction” treadmill rather than a hedonic treadmill; people who have more bad experiences may declare themselves satisfied with a poorer distribution of good/bad experiences but the number and nature of good/bad experiences do not alter as a result of past experience). He argued further in this regard that if hedonic adaptation were thoroughgoing—rather than hedonic reporting being labile—we would not observe, as we do, that certain experiences (like cutting oneself shaving) are always unpleasant, no matter how often they are repeated. But if the most significant impact of adaptation is to broaden the indifference band (and to broaden it systematically to include more events of the “familiarly” hedonically charged sort), then the purported distinction between the old literature’s “reporting errors” and the new literature’s “unmediated” responses largely disappears.

Kahneman’s theoretical argument is, in my view, quite unpersuasive. Even if good/bad judgments are bipolar and nonrelativistic, there is no reason to believe that the “neutral” (zero point) area does not increase in size. (Thus, it may be the case that those who have experienced lots of misery find a wide
range of once-miserable experiences too ordinary to condemn or avoid, or that those who have done lots of wonderful things become jaded in the sense that a far larger range of experiences are neither aversive nor attractive but neutral). If this is true, the "hedonic treadmill" is real, rather than a satisfaction-reporting treadmill: it will simply be manifest as a wider zero point for once-pleasurable experiences. Kahneman, Experienced Utility, supra note 5, 11-12.

My tentative sense is that this problem is related as well to the distinction between "shifting adaptation levels" and "desensitization." See, in this regard, Shane Frederick & George Loewenstein, Hedonic Adaptation, in WELL-BEING, FOUNDATIONS, supra note 5, at 302, 303 ("It is important to distinguish between adaptive processes that diminish subjective intensity by altering the stimulus level that is experienced as neutral (shifting adaptation levels) and adaptive processes that diminish the subjective intensity of the stimulus generally (desensitization).”). If you just adjust adaptation levels, you should still be able to sense distinctions between stimuli. Thus, assume you go to prison: after a time, confinement in a seven-foot cell may become less bothersome, but the distinction between a seven and nine-foot cell might actually be heightened (so you would have experienced adaptation but not desensitization). My intuition—though it is one I have little confidence in—is that desensitization among some range of the options that one actually most typically experiences might be one way of describing a radical increase in the no-response area.

As I noted, Kahneman believes not only that desensitization is not a very important part of the hedonic treadmill process, but hypothesized in his published writing that people do not really adjust adaptation levels either. Rather, the subject simply changes how he reports the "summary" of any given level of actual satisfaction. Note, though, that there are some indicators that what psychologists have perceived as adaptation is not simply a reporting/memory problem, but is manifest in non-communicative behavior. See E. Krupat, Context as a Determinant of Perceived Threat: The Role of Prior Experience, 29 J. PERSONALITY & SOC. PSYCHOL. 731 (1974) (finding that prior exposures to threats not only lowers subjects' reports of how dangerous the situation they are in is but reduces galvanic skin conductance); Dar Reuven, Dan Ariely & Hanan Frank, The Effect of Past Injury on Pain Threshold and Tolerance, 60 PAIN 189 (1995) (finding that veterans who had suffered more serious injuries in the past both waited longer before reporting severe pain when exposed to 48 degree Celsius water, and, behaviorally, terminated the test by withdrawing their hands form the water later). But there is contrary evidence as well. See, e.g., Paul Paulus, Garvin McCain & Verne Cox, A Note on the Use of Prisons as Environments for Investigations of Crowding, 1 BULL. PSYCHONOMIC SOC'Y 427 (1973) (finding that prisoners do not report higher levels of feeling crowded when density rises, implying hedonic adaptation, but they exhibit physical symptoms of stress, like palm sweating more, implying that they have not adapted).
have suggested, that a certain degree of duration neglect may be appropriate because persistent pain in a (conceptually construed) single domain is less bothersome than equal or even lesser experienced pain that arises from (what the subject conceives of as) multiple sources, then the ecological momentary assessment will always "overstate" the degree of misery that the chronically ill live with and understate the degree of misery that those who lack the internal and external resources (power) to protect themselves against multiple affronts suffer. Now, of course, there is a more straightforwardly political translation of the weak perfectionist concern there: using the ecological momentary assessment as an elicitation method may implicitly under-value the importance of "power" (the capacity to resist bad events) in "welfare." What is ultimately critical though is that no matter what elicitation technique we use, we are measuring something that could aptly be thought of as a form of welfare: the choice to privilege one or the other elicitation mode reflects nothing more than our own (weak perfectionist) beliefs about which form of "welfare" people might value is more expressive of desirable human aspirations.

C. Is "Happiness" Straightforward and Unambiguous Because its Biological Role is Simple?

A possible rejoinder to the claims I have made that happiness is too multi-faceted to capture using any single measurement mechanism derives from one interpretation of the literature on the biological role of pain and pleasure. In this view, both happiness/pleasure and unhappiness/pain

---

80. There is some considerable congruence here between the contrast between feeling empowered and feeling a good balance of painful and pleasurable moments that I am drawing and the contrast some (but not all) psychologists draw between "mood" (the felt capacity to meet challenges) and "emotion" (momentary affect.) See supra note 55.

81. It will also arguably over-weight aspects of life that take up more of the day, even though these domains do not weigh quite so much on reflective understanding of one's life. Thus, someone with a ten hour a day cruddy job might press the "get me out of here" response button on the Palm Pilot far more hours a day than someone with a benign job and a cruddy marriage (unless the cruddy marriage actually weighs on his hedonic state throughout the day, which it might or might not). It is not clear, though, that the self-reflection of the person who thinks he is satisfied with his life overall, given the relative importance of work and family, is entitled to no weight.
really are far more unsubtle, straightforward phenomena than I have implied because each serves a particular narrow biological role: either to induce change—NB for those of you not good at life management, pain's the one that should be doing that—or persistence (in the case of pleasure). However, the most plausible reading of the biological literature is that the roles of pleasure and pain are hardly transparent, though I acknowledge that there are two facially colorable readings of the literature that at least suggest that two of the problems to which I have adverted can be overcome.\(^8^2\)

First, it is possible to argue that people are happy when their status is (objectively) good and unhappy when it is (objectively) in need of change. This seems to permit us to answer one question I have raised about the propriety of welfarism.\(^8^3\) Welfarists, armed with the knowledge that people's happiness reflects their "proper functioning" can then defend themselves *either* by disclaiming any (openly perfectionist) interest in insuring that people "flourish" (noting that concepts of flourishing are too controversial and unshared to form the basis of ethics) *or* by arguing that people who are happy *are* flourishing, by biological definition, given the most plausible accounts of what it might mean to flourish. Second, it is possible to argue, more modestly, that even if we cannot say that those who are happy are inevitably functioning "well" in some uncontroversial way, we can identify happiness as a stable phenomenon—should we decide to care about it—because it is the unambiguous state that people are in when they are approaching, rather than avoiding, the perceived salient features of their environment. I take up each point in turn.

---

82. The obvious problem in getting to that point is that I am not a biologist, I cannot even do a decent job critically evaluating biology, and cannot stand it when social science types believe they can draw important lessons from science that they barely comprehend. Nonetheless, since I think it is tempting to try to use biology to "rescue" hedonic welfarism in one of two ways, but that neither ultimately works, I feel obliged to comment a bit on "hedonic biology."

83. This question is the one I posed, *see supra* note 1, about whether third party observers *should* care about how well-off people perceived themselves to be, rather than any other teleological end-state. One version of the argument from biology is that it gives more content to the dream that all reasonable perfectionists seek: to insure that people are "flourishing."
1. *Does Happiness Imply Functionality?* Assume, solely for argument's sake, two quite controversial propositions: First, assume (descriptively) that it is the case that people are happy when doing things that help them (directly or indirectly) survive or procreate. Assume, then, (normatively) that at least some perfectionists interested in human flourishing would argue that people are flourishing to the degree that they are successfully fulfilling their most basic biological role (to pass on their genes). What is not at all plain is that it would follow that subjects who are happier on balance are more "successful" in meeting the posited end (gene-line propagation). They might, for instance, face fewer (painful) threats to their capacity to meet this end, but might cope so poorly with the ones they do face that they fail in their "evolutionary" task.

In fact, it would seem that hedonic psychologists who took the tack of arguing that happy people are those that are functioning well would need to confront the "insight" (of great and obvious appeal to the gloomy, of somewhat less certain evolutionary biological pedigree) that people are evolutionarily biologically predisposed to be unhappy on balance. The basic intuition behind this view is that our

---

84. Naturally, these propositions are each enormously controversial. I discuss the problems with the descriptive proposition in section 2, below. The normative proposition is obviously hardly self-evident either: some class of observing perfectionists would plainly think it far from obvious that someone is flourishing in the way that most expresses what is special about human nature if all he has been is especially successful at passing along genes.

Here's a parallel point, drawing on the Scitovsky work I adverted to earlier on the "comfort/pleasure" trade-off: one "reason" we may (arguably) seek comfort, even though it interferes with "pleasure," is that for most of our evolutionary history, it was necessary to seek comfort, and there was no reason to anticipate the possibility that we would satisfy our need for comfort so readily that doing so would induce boredom. But does the fact that there might be a sense in which there is an atypically strong biological basis for (one sort) of satisfaction tell us anything at all useful about the normative status of that satisfaction? What survival-oriented instincts direct us towards and what either perfectionists or utilitarian evaluators might think we should try to maximize need have no overlap. *See generally Scitovsky, supra* note 54.


Pleasure is always contingent upon change and disappears with continuous satisfaction. Pain may persist under persisting adverse conditions. One gets used to the events that, earlier, delighted and caused joy; one does not get used to continuous harassment or humiliation. . . . The law of hedonic asymmetry is a stern and bitter law. It seems almost a necessary one, considering its roots, which,
only real biological need to which approach and avoidance are directly relevant is the need to do what it takes to get old enough to reproduce; in order to do that, we need to develop strong feelings of anxiety and pain (since the feelings must be so compelling that it seems imperative to do something to avoid threats to our survival). Happiness, in contrast, can be fleeting and not very intense and noticeable, since all it signals is that we need not do much. Note that this argument, like most such socio-biological "just-so" stories is not overwhelmingly compelling: it might be, for instance, that those who are—or appear to be—happy are better able to attract mates, or might, as a result of—even unwarranted—optimism, invest more in their children, increasing the odds that their genetic line persists. If one couples the observation that there may be biological reasons to be more sensitive to displeasure than pleasure with an observation that may be the core insight

theoretically, are so obvious. Emotions exist for the sake of signaling states of the world that have to be responded to or that no longer need response and action. Once the 'no more action needed' signal has sounded, the signaling system can be switched off; there is no further need for it. That the net quality of life, by consequence, tends to be negative is an unfortunate result. It shows the human mind to have been made not for happiness, but for instantiating the blind biological laws of survival.

Id. at 353-54.

But note in this regard the competing (and at least equally plausible) idea that it is not so much persistent misery that is functional as something more like hedonic adaptation. See Frederick & Loewenstein, supra note 79, at 302-03 ("Because the persistence of an aversive state is an indication that it cannot be changed, hedonic adaptation may prevent the continued expenditure of energy in futile attempts to change the unchangeable and redirect motivation to changes that can be made."). In one view then, hedonic adaptation makes us sensitive to small, incremental local changes that are now most likely to be action-relevant. I may dread (roughly equally) confinement in a seven or nine foot cell, but if I "adapt," by virtue of a prison sentence, so that seven is closer to neutral and nine on the positive end, I may take steps to get the nine foot cell reward. Moreover, the persistence of strong negative emotional states (e.g., fear) causes health problems so they are plainly dysfunctional at some point.

86. I have got to confess that I cannot recall ever meeting a socio-biological argument that seems better than cute. I am partly driven in my critique of such "species survival" stories by the brilliant observation of the late Amos Tversky: spotting a truly mangy looking mutt, panting hideously and slobbering away as it skulked in the foothills near the Stanford dish, he turned to a graduate student friend and said, "Always remember that evolution is a floor, not a ceiling."
merging existential philosophy and prospect theory—of course, there is gain/loss asymmetry, dimwit, what is any good thing you could possibly name compared to death and the dread of death?—then there is reason to doubt that (overall) happiness signals that one is in "the most functional biological state."

2. Is Happiness Nonetheless a Functionally Defined State? Even if those arguing that happiness is simple and straightforward concede that happier people are not inevitably in globally superior shape, they may still argue, in two distinct ways, that happy feelings can be identified straightforwardly because they have a particular role. But there is reason to doubt as well either that happiness is that state needed to guide the organism to "high functioning" outcomes or even to doubt the more modest claim that it is simply that state that motivates a certain kind of action (approach).

In this first sense, observing that X is happy (overall) does not precisely tell us that he is functioning well (generally), but tells us, simply, that that he is doing things that will prospectively be good for him. Being happy (unhappy), on balance, may or may not be good, but being happy (unhappy) is a very particular signaling state—the state signaling that one is in the midst of a biologically beneficial (problematic) activity.

Even that statement seems exaggerated. Indeed, it appears that people have certain innate tastes that are readily biologically comprehensible. The sweet and fatty foods that we are predisposed to enjoy have life-sustaining calories, while most naturally occurring dysphorically bitter tasting substances are poisonous. But extremely sour tastes—to which we are also apparently naturally averse—scarcey occurred in nature and seem to have no positive or negative survival valence. The "role" of taste and distaste in our biological lives may thus often be muted. There is thus no general reason to believe that we are serving our biological aims better when we are happy rather than unhappy—being unhappy because we are tasting bitter poison may indeed indicate that we are at (biological) risk,

but being unhappy that we are tasting sour lemon juice indicates no such thing.

Moreover, because both aversion and attraction mechanisms can readily be "captured," or subject to odd interferences, we should expect there to be an even greater disconnect between either success or failure in performing "survival" tasks and hedonic states. Animals may "feel" happy, though they are not performing pro-survival tasks because the mechanisms that make them happy or unhappy when performing the tasks are "sabotaged." Happiness does not directly lead organisms to choose well; it is merely the state that most typically accompanies pro-survival choices. But even when the creature acts "appropriately," it may not be happy, and it may not be

88. Food disgust may conceivably play an important evolutionary role—arguably, if people were not disgusted by some food, they would eat things that would kill them. (Though as I note later, they might be able to avoid bad foods without disliking them in the usual sense). But the mechanisms that permit evolutionarily necessary disgust to function might be co-opted to insure that people will feel identical disgust when confronted with foods that are thought of as religiously impermissible, even when those foods are perfectly safe. Id. at 110.

89. This observation should lead us to raise, once again, the philosophical question about why a third party observer should care at all about maximizing welfare. It is not clear that we should care about happiness if the only reason we purport to care about it is that it (generally) seems to occur when an organism is engaged in "appropriate" (gene-replicating) activities. If that were the case, it would seem we would judge the organism to be successful when and only when it is engaged in "good" tasks ("objectively" identified pro-survival tasks) and view happiness (at best) as an over and under-inclusive surrogate for doing such good tasks.

90. The biological evidence on this is very hard for me to sort through with any degree of confidence, but here is the way I read a series of older experiments on maternal behavior in rats. Rats are one of those species that raise only their own young; female rats not only do not care for the young generally, but may well eat them. Let's assume that experiments establish pretty convincingly that the non-lactating adult females are in some sense "unhappy" in the presence of baby rats because the babies are essentially aversive-smelling to most adult female rats. New (lactating) mother rats are both flooded with the standard chemical source of "happiness" (oxytocin) and suppress the aversive smell. Thus, the happiness-is-the-emotion-the-organismFeels-when-acting-functionally story is that in an animal group in which it appears optimal (in the pass-along-genes sense) for mothers alone to care for the young, recent mothers are happy caring for the young, and other females are not. But "happiness" is more directly mediated by physiological mechanisms that can readily mis-fire or be "hijacked;" if such hijacking occurs, there is no reason that creatures will be happy or unhappy at "appropriate" times.
moved by "mere happiness" to choose to do something advantageous.

The deeper problem in relationship to people is that the "capture" or "sabotage" problem may well not be a sideshow, created by experimental manipulation (like the removal of a rat's sense of smell, designed to "trick" non-lactating female rats into finding care of other rat pups less aversive). The direct survival value of the bulk of human tasks is really questionable, and we may have our pleasure and pain centers consistently activated for reasons of little biological moment. For instance, people learn to enjoy certain (innately) unpleasant (bitter) tastes (e.g., coffee) by associating them for a period with (innately) desired tastes (e.g., cream and sugar); figuring out what the biological gain in doing this will not prove easy unless, in some sense, we seek (some form of) pleasure for its own sake, rather than to direct us to seek some further, biologically "sensible" end. Similarly, people may experience disgust (and thus act as aversively as they would act if in direct

If you render the non-mother female rats anosmic (e.g., remove the sense of smell, typically with zinc sulfate), they get far more willing (happier?) to be around the babies (though still less than the lactating recent moms). However, merely injecting them with oxytocin (to make them happy) in the baby's presence is not sufficient, because one has not sabotaged all of the mechanisms that create approach and aversion. The basic findings are reported in Marianne Z. Wamboldt & Thomas R. Insel, The Ability of Oxytocin to Induce Short Latency Maternal Behavior is Dependent on Peripheral Anosmia, 101 BEHAV. NEUROSCIENCE 439 (1987) and Alison S. Fleming & Jay S. Rosenblatt, Olfactory Regulation of Maternal Behavior in Rats, 86 J. COMPARATIVE & PHYSIOLOGICAL PSYCHOL. 233 (1974). At the same time, removing lactating mothers' sense of smell seems to interfere with their performing maternal tasks even though they are still atypically oxytocin receptive. Thus (to try to put the point in philosophical perspective), maternal tasks do not (inevitably) make the lactating rats happy in any direct sense (though they are equally "fulfilling" of biological imperatives) nor does feeling happy around babies inexorably induce care-taking, but rather the "inclinations" to perform the tasks flow through a series of pathways that either function in a typical fashion or do not. Happiness (and lack of aversion) usually accompany species-appropriate, gene-line preserving tasks, but they need not. Moreover, while happiness might be thought to be sufficient to channel behavior (even if artificially associated with a task), it need not be sufficient unless other behavior-determinants (like smell aversions) are toyed with as well. See also Irwin Benuck & Frank A. Rowe, Centrally and Peripherally Induced Anosmia: Influences on Maternal Behavior in Lactating Female Rats, 14 PHYSIOLOGY & BEHAVIOR 439 (1975).

91. See Debra A. Zellner et al., Conditioned Enhancement of Human's Liking for Flavors by Pairing With Sweetness, 14 LEARNING & MOTIVATION 338 (1983).
sensory pain) that was never conditioned on any direct sensory reaction: aversion reactions are “captured” by non-sensory, meaning-infused aspects of the person. Thus, for instance, if one (culturally and cognitively) construes food as coming from a “forbidden” origin (think of eating something one thinks is incredibly “gross”), one will disdain it as much as he would disdain something he experienced in a sensory fashion as unpleasant. For instance, the same smells are experienced as quite distinct if subjects are told they are cheese rather than cat feces. People do not seem to act (much of the time) based on a culturally unmediated set of hedonic reactions; it is, once more, hard to see the direct biological gain in having aversion reactions “captured” by culture.

Can we nonetheless at least define happiness and unhappiness as those feelings that motivate distinct sorts of behavioral reactions (approach and avoidance), even if being happy on balance is not in some strong sense itself an optimal biological state or being happy at a particular moment is not a signal that one is engaged in a task that will help us approach such a state? I still think not for a variety of reasons also suggested by a review of the biological literature on pleasure and pain.

First, it is possible (for people and animals alike) to learn (through simple conditioned learning, requiring only some degree of negative experience on some occasions) to avoid things without (more globally) disliking them so that it is not clear that the “role” of disliking is to fill a necessary steering role. Again, here is the broader conceptual point: if it is true that it is not obviously necessary to “dislike” things in order to avoid them, it may not be possible to define unhappiness/disliking as that state that exists to direct us to shift behavior/avoid the immediately present

---

92. For a discussion on this subject, see Rozin, supra note 87, at 117-118.
93. Some may claim that pain is an especially compelling motivator—I think this is the mainstream biological answer to Yossarian’s question in Joseph Heller’s Catch-22 about the failings of the Supreme Being: “Why in the world did He ever create pain?... Why couldn’t He have used a doorbell instead or one of his celestial choirs? Or a system of blue-and-red neon tubes right in the middle of each person’s forehead. Any jukebox manufacturer worth his salt could have done that. Why couldn’t He?” JOSEPH HELLER, CATCH-22 189 (paperback ed. 1996). The mainstream answer does not appear quite adequate to account for all the levels of pain, and all of the levels and varieties of dislike that we feel.
situation. In fact, a wonderful, curious finding in food research is that we invariably come to dislike food that has made us nauseous but (learn to?) avoid, without disliking, food that has given us allergic reactions or upper gut pain. More generally, it seems possible to activate the "choosing" or activity-directing aspects of the brain independent of the "evaluative" aspects (even though they most typically mutually interact); thus, choosing and liking are at core simply not the same phenomenon. Mix water with sub-clinical doses of cocaine for recovering drug addicts and they will choose it over a saline solution at wildly disproportionate rates, but not only do they claim that it gives them absolutely no pleasure, they have no cardiovascular response to the drug. One can get similar dissociations between "wanting" and "liking" on the negative ("avoiding" and "disliking") side. Thus, for instance, arachnophobic subjects given naltrexone, a drug that mimics naturally occurring proteins bound to receptor cells, thus blocking endorphin and enkephalin neurotransmitters that might otherwise reduce anxiety, refused to approach a spider at an earlier stage of a string of "approach tasks" (look at the spider through a glass,

---

94. Marcia Pelchat & Paul Rozin, The Special Role of Nausea in the Acquisition of Food Dislikes by Humans, 3 APPETITE 34, 348 (1982).

95. See Marian W. Fischman, Relationship Between Self-Reported Drug Effects and Their Reinforcing Effects: Studies with Stimulant Drugs, 92 NIDA RES. MONOGRAPHS 211, 220 (1989) ("Subjects reliably chose 8 mg cocaine over 4 mg, and both doses were reliably chosen over saline. Neither self-report data nor cardiovascular measures indicated any differential effects of these lower intravenous doses."); Marian W. Fishman & Richard W. Foltin, Self-Administration of Cocaine by Humans: A Laboratory Perspective, in COCAINE: SCIENTIFIC AND SOCIAL DIMENSIONS 165, 169 (Gregory R. Buck & Julie Whelan eds., 1992) (finding that drug users selecting low doses of cocaine while being administered desipramine continue to select the drug over saline solutions though it seriously modified their positive subjective reactions to the drug; thus, verbal reports of euphoria or positive drug effects and drug self-administration dissociate from each other under a number of conditions). See also R.J. Lamb et al., The Reinforcing and Subjective Effects of Morphine in Post-Addicts: A Dose-Response Study, 259 J. PHARMACOLOGY & EXPERIMENTAL THERAPEUTICS 1165 (1991) (finding that though subjects persistently chose the lowest dosage of morphine over a placebo, subjects did not report distinct subjective effects from the low dosage of morphine; thus, the reinforcing effects of morphine can be manifest without the drug having subjective effects. Only 38% of users even knew they had received the 3.75 mg dose of morphine they had selected. Moreover, there was no correspondence between the degree to which particular patients reported that they liked low dosages of morphine and their tendency to self-administer the drug the next time they were given an opportunity to do so).
touch the spider with a pencil, touch the spider...), but reported no higher levels of fear of the spider than those receiving a placebo. Again, the message from the vantage point of "hedonic philosophy" is more straightforward than the underlying neuroscience: one cannot simply identify "things we are driven to choose" with "things that we like" (even prior to reaching the conventional complications in preference-utilitarianism about misinformation, regret, ambivalence and the like). Most typically, but not always, liking is the emotional state that drives us to seek, but this is not physiologically—and hence definitionally—necessary.

Similarly, people (unlike animals) do not always choose to reject what they do not like. Humans appear to be unique (among the species that we study) in deciding to embrace things we (in some sense) continue to "dislike." It would doubtless be interesting to understand more than I do to do so about why we most typically seem when our reactions are mediated through social learning rather than when we act as isolated individuals—though one would have to understand better than I do non-Freudian psychological views of masochism to test the proposition that most instances of approaching the painful are socially accepted or generated. Some psychologists argue that it is a result of some complex—and less socially mediated—tendency to take pleasure in things that ultimately prove harmless though they once appeared dangerous. In this regard, it is important to think about our attraction to constrained risk-taking generally (e.g., think of roller coasters, scary movies, faculty meetings). Thus, to draw on the food literature again, people are naturally initially averse to (hot) peppers but in a large number of cultures, they "train"

96. See Arnoud Arntz, Endorphins Stimulate Approach Behaviour, But do not Reduce Subjective Fear: A Pilot Study, 31 BEHAV. RES. & THERAPY 403 (1993). I am afraid that this is one of many places in which my extremely limited knowledge of the underlying biology haunts me. I suspect, but am not really at all sure, that these sorts of findings are connected to findings that people can (at least under certain atypical circumstances) have "knowledge" without "awareness" of that knowledge. See, e.g., Daniel Tranel & Antonio R. Damasio, Knowledge Without Awareness: An Autonomic Index of Facial Recognition by Prosopagnosics, 228 SCIENCE 1453 (1985) (stating that patients with prosopagnosia could not consciously recognize familiar faces but their electrodermal skin responses were elevated when shown pictures of people they once consciously knew).

97. See generally D. E. BERLYNE, AESTHETICS AND PSYCHOBIOLOGY (Kenneth Mac Corquodale et al. eds., 1971).
themselves to like them (usually between the age of 4 and 7). This is true even though those who love the peppers experience precisely the same neural signals from their mouths (harmless irritation from exposure to capsaicins) that they experienced when still finding them unambiguously aversive.  

II. PREFERENCE UTILITARIANISM

A. Preliminary, Peripheral Problems: Maximizing Preference Realization and the Problem of Non-Self-Regarding Preferences

I need to pay a certain degree of attention to two thorny, preliminary problems in preference utilitarianism, even though each of them is, fundamentally, from the perspective of "hedonics," a side issue. The point of this brief sub-section, in essence, is to limit the domain of the more basic discussion of preference utilitarianism in the next sub-section to the (relatively) simple case in which some subject S expresses a preference between some self-regarding end-state X and some other self-regarding end-state Y, and we therefore conclude that he is "happier" if he gets X than Y. In order to restrict the discussion to that "simple" case, though, I must first discuss problems that we would encounter if we tried to use preference utilitarian methods to evaluate how subjectively well off a person was with her life on the whole, rather than merely evaluating whether her subjective welfare improved or declined when one of two events in a restricted choice set transpired. It is plainly far more difficult to figure out the degree to which a person's elaborate and multi-dimensional preference set is satisfied than to determine whether the person received a particular preferred alternative to a less-preferred alternative. Second, it is also not easy to determine the degree to which we should assume subjective well-being derives from preference-satisfaction generally, or only the satisfaction of some sub-set of preferences usually dubbed "self-regarding" preferences. In this regard, I must briefly

98. See Paul Rozin, Getting to Like the Burn of Chili Pepper: Biological, Psychological, and Cultural Perspectives, in CHEMICAL SENSES: IRRITATION 231, 248 (B. C. Green et al. eds., 1990).
note some of the difficulties involved in defining self-regarding preferences, and also deal, albeit cursorily, with the question of whether non-self-regarding preferences are preferences merely to pursue certain ends or whether meeting these preferences matters as well.

Preference-utilitarianism is easiest to implement when it restricts its domain to pair-wise ordinal comparisons; it is most straightforward to comprehend the statement that the subject S is “better off” consuming X than Y if she would prefer X to Y. (Setting aside, for the moment, questions I return to about whether the preference was formed with adequate prudence and information). But it is not even readily comprehensible to say that S is better off if she gets “more of her preferences” satisfied or “more of her desires met” or that S is better off than T if she gets “more of her desires met” than T gets of his. Plainly, it would be nonsensical to think that if S happened to express her desires as a long list of discrete wishes, most of which got met, she would be happier than if she would have combined many of the wishes into a unified end-state description and noted that she was frustrated in her desire to meet the restated goal. Thus, if A desires X, Y and Z and can get X and B desires P, Q, and R and can get P and Q, it is not at all clear that a preference utilitarian would say that B is better off than A (or, correlatively, that we have maximized utility if B rather than A gets what she can get); nor would we even say that A is happier if she gets Y and Z rather than X, unless she has got a preference between those two bundles.

This particular problem in preference utilitarian thought is soluble (in theory if not in practice). Harsanyi,

99. In Ronald Dworkin’s renowned work rejecting distributive principles that attend to the distribution of welfare, he seems to attribute to preference utilitarians the rather incomprehensible view I deride in the text that they seek to maximize the extent to which subjects satisfy as many preferences as possible. See Ronald Dworkin, What Is Equality? Part I: Equality of Welfare, 10 PHIL. & PUB. AFF. 185, 204-220 (1981). Dworkin seemingly finds the following statement intelligible, “people are as nearly equal as distribution can make them in the degree to which each person’s preferences about his own life and circumstances are fulfilled.” Id. at 204-05. I take it that this same confusion leads him to discuss whether easily-satisfied people should be considered better-off, in a preference-utilitarian sense, than people with greater ambitions, because they (presumably) have “more” of their preferences satisfied, without regard to whether they “meta-prefer” the bundle that is in fact satisfied. Id. at 207-16.
among others, reduces all preferences to pair-wise choices (of increasingly complex bundles that include one's reaction to frustration of concrete preferences or risk), and deals with the interpersonal utility comparison quandary by imagining a neutral agent who chooses between the totality of one agent X's life circumstances (first assuming he has that agent X's tastes) and another agent Y's whole life (assuming he has Y's tastes). Within such a scheme, X is better off than Y if the neutral agent would choose to be in X's position (just as S herself is better off in position 1 than 2 if she would prefer 1, overall). Thus, questions about whether we characterize X as having a high percentage of her (highly disaggregated) preferences satisfied or a low percentage of her more aggregated ones become trivial and irrelevant. Similarly, questions about whether X appears more satisfied solely because she demands less will disappear if a neutral agent with X's meta-tastes would not choose X's relatively satiated, low-ambition life.

It is not clear, as well, whether preference utilitarians should restrict the domain of preference utilitarianism to choice-pairs with immediate effects on the selfish well-being of the subject. Even if subject S is better off in some conventionally recognizable sense if he gets the vanilla ice cream he prefers to the chocolate he disfavors, it is not clear whether he is “better off” (in the same sense?) if we collectively follow the policies he favors in relationship to endangered species. (In distributive ethics terms, is X “better off” than Y if we follow her preferred political policies—when she cares intensely about politics—but she gets far fewer material goods?).


101. For some of the many good (and a good deal fuller) discussions of the issues in these next few paragraphs, see, e.g., GRIFIN, WELL-BEING, supra note 2, at 13-14, 19-20; JOEL FEINBERG, HARM TO OTHERS 83-89 (1984); PARFIT, REASONS AND PERSONS, supra note 67, at 494-95; Shelley Kagan, The Limits of Well-Being, 9 SOC. PHIL. & POLY 2 (1992); D.W. Haslett, What is Utility?, 6 ECON. & PHIL. 65, 80-82 (1990); Ernest Partridge, Posthumous Interests and Posthumous Respect, 91 ETHICS 243 (1981); David Sobel, Well-Being as the Object of Moral Consideration, 14 ECON. & PHIL. 249 (1998).

102. Note that there may be reasons that we ought not to design political policies so as to satisfy the non-self-regarding preferences of citizens, even if we thought people were made better off when such preferences were satisfied in precisely the same way they would be made better off if other preferences were met. Dworkin, not atypically, argues that moral/political principles must
Non-self-regarding tastes may quite regularly be satisfied without the people whose tastes have been met even knowing they have been satisfied: connecting the unknown satisfaction of desires to well-being is plainly troublesome. Still, people could rationally choose to seek some end X (e.g., the restoration of an endangered species to non-endangered status), perfectly aware that they might never learn whether X occurred, rather than some self-regarding end Y whose fulfillment (or frustration) would be unambiguous. Someone might unambiguously believe her life to have been “better” (more satisfactorily) lived pursuing certain ends. It would nonetheless be peculiar to think she was indifferent to whether the ends were actually attained; thus, it would be incomplete and inaccurate to say that she simply preferred the activity of pursuing X to the activity of pursuing Y. Moreover, people could readily have powerful impersonal ends that they do not actively pursue (e.g., rooting for a sports team to win a championship).

It would be odd to say that a person could not be made happier by learning his favorite team won the championship than he might have been made by the fulfillment of some self-regarding desire (e.g., to eat at an excellent restaurant) just because the desire (for his team to win) could have been met without his knowledge or after he was dead, or because he did nothing to bring the result about. And yet it would also seem to be troublesome to translate all (potentially) impersonal desires into personal ones. While it might be accurate to translate, “I wish the

precede tastes: “[A]n inegalitarian political system does not become just because everyone wrongly believes it to be.” DWORKIN, supra note 99, at 201. What is more relevant to the discussion in the text though is that Dworkin further believes more broadly that we should not attend to the satisfaction of any “impersonal preferences” (e.g. a desire that life be discovered on Mars) though, plainly, attending to such preferences need not compromise the formation of just policy. Id. at 201-04.

103. In a parallel vein, Cohen criticizes what he considers Sen’s unduly “athletic” (or activity-centered) conception of “capabilities.” Sen’s ideal metric for distributive ethics focuses heavily on the distribution of such capabilities, but Cohen notes that we might get a good deal out of things we do not do but that happen to us. See G.A. Cohen, Equality of What? On Welfare, Goods, and Capabilities, in The Quality of Life 9, 23-28 (Martha Nussbaum & Amartya Sen eds., 1993).

Griffin argues that an exclusive focus on “project pursuit” reflects a masculine bias; women, he argues, are more likely to appreciate and value things that happen to us (like being enmeshed in significant relationships). See WELL-BEING, supra note 2, at 22.
Dodgers would win the World Series” into “I wish I would have the experience of watching or hearing about the Dodgers winning the World Series,” it would not be accurate to translate, “I wish that ethnic violence in the former Yugoslavia would abate” into “I wish I would know for sure that such violence had abated.” Many subjects would argue that the impact of the end of the ethnic violence on their happiness is, at core, beside the point. Learning of the end of the violence might be more commensurate with sensory experiences than the wish to end the violence is commensurate with a wish for a sensory experience even if the sports fan’s “impersonal” wish might be fully commensurate with a sensory wish because its fulfillment outside the realm of experience would be valueless.

B. The Basic Problem: Laundering Manifest Preferences

Even if we put these thorny issues aside, we must consider whether some subject S is better off if he achieves some self-regarding end-state X rather than Y if he prefers that state X to Y. The standard answer is that he does so only if he were, at the very least, rational, to at least in the sense that his preferences obeyed certain consistency axioms, and that he is adequately informed about the

104. Some writers identify “rationality” with something more like what I describe as “prudence” in the text. For instance, Richard Brandt argued that one’s desires are not truly rational unless they would survive “cognitive psychotherapy.” See Richard B. Brandt, A Theory of the Good and the Right 113 (1979).

105. For those readers (especially those schooled in economics) accustomed to thinking not about the consistency of underlying preferences, but of observed choices, it is important to note that it is really preferences, not choices—or what mainstream economists call revealed preferences—that must be consistent. If S chooses X over Y and then chooses Y over X, what looks like inconsistency could be a function of changing tastes or circumstances (e.g., he has become relatively satiated with X having chosen and consumed it). Since multiple choices cannot be made instantaneously, the theory of revealed preference is not empirically falsifiable. Rather it relies on precisely the sort of intuitions about internal psychological states that it seeks to avoid by reference to behavior. For classic formulations of revealed preference theory, see Paul A. Samuelson, A Note on the Pure Theory of Consumer’s Behavior, 5 Economica 353 (1938) and I. M. D. Little, A Reformulation of the Theory of Consumer’s Behavior, 1 Oxford Econ. Papers 90 (1949).
qualities of X and Y and is adequately prudent in making the choice. There are a host of differentially stringent accounts of when a person is either adequately informed or adequately prudent, but the variations, for my purposes, are essentially beside the point.

There is an important ambiguity buried in all variants of the requirement that the subject be fully informed. Must he simply be more fully informed about the actual, precise physical nature of the goods he prefers or rejects, or must he have more information about the causal relationship between satisfying his desires and his ultimate well-being? In the former case, we do not need to "correct" his desires, fully explicated. We merely need to correct how he has translated these desires into concrete, particularistic practice. Subject S already knows that he prefers good wine.

This point is made quite powerfully by Sen:

Preferring x to y is inconsistent with preferring y to x, but if it is asserted that choice has nothing to do with preference, then choosing x rather than y in one case and y rather x in another need not necessarily be at all inconsistent. What makes them look inconsistent is precisely the peep into the head of the consumer, the avoidance of which is alleged to be the aim of the revealed preference approach. It could, however, be argued that what was at issue was not really whether the axiom of revealed preference represented a requirement of consistency, but whether as a hypothesis it was empirically verified. This line would not take one very far either. . . . Comparisons have to be made within a fairly short time to avoid taste change, but the time elapsed must also be sufficiently long so that the mutton purchased last time is not still in the larder, making the choices non-comparable. . . . In fact, the concept of taste change is itself a preference-based notion. . . .


106. Some theorists, like Griffin, seem to demand no more than that the person expressing a desire be formed an appropriate "appreciation of the nature of its object." Well-Being, supra note 2, at 314-15 n.19; James Griffin, Against the Taste Model, in Interpersonal Comparisons of Well-Being 49 (Jon Elster & John E. Roemer eds., 1991). Others seem to demand that the subject possess and process a good deal more information before choices are adequately "corrected." See, e.g., D. W. Haslett, What is Utility?, 6 Econ. & Phil. 65, 72-74; John C. Harsanyi, Morality and the Theory of Rational Behavior, in Utilitarianism and Beyond, supra note 45, at 55.

107. See Arthur Ripstein, Preference, in Value, Welfare, and Morality 93, 103 (R.G. Frey & Christopher W. Morris eds., 1993) ("In deliberating about how best to satisfy my ends, extra information is always useful, subject only to time constraints. Yet if that is all that information is doing, it would seem that I already know what is good for me, and only be concerned about how to get it. In contrast, if I am reflecting about what ends matter to me, the role of extra information is less clear.").
to water to poison, but when he looks at the liquid in glass A, he mistakenly believes it is good wine, not poison. His (long-list) desires, in that case, need not be “corrected;” he must simply translate desire into choice differently. In some sense, it would be odd to think he “desired” the liquid in the glass (whatever it turned out to be)—true desires are simply more abstract and end-state oriented than that. Or, to use Sumner’s example, S desires a Caribbean vacation, unaware that taking it will put him in the path of a dangerous hurricane. If we knew the full structure of his preferences (including his priorities and trade-offs among them), we would know he would reject the concrete vacation plan if he knew the hurricane would follow its actual ultimate path.

108. Clearly, though, it is the case that economists frequently move (carelessly) from the political proposition that it might well be (politically and pragmatically) appropriate to allow persons freedom to live in accord with their own decisions to the clearly false view that people invariably choose what is best for them. This point was clearly, if not originally, pointed out in John Broome, Choice and Value in Economics, 30 OXFORD ECON. PAPERS 313 (1978). It is also the subject of one of the first pieces I published as a legal academic. See Mark Kelman, Choice and Utility, 1979 Wis. L. Rev. 769.

109. SUMNER, supra note 3, at 131.

110. I have no doubt that the line between these two forms of information-correction is less hard and fast than I imply in the text. Assume, for instance, that one thinks a patient is “under-informed” about how painful a medical procedure or life event will be. One could readily say she simply misapprehends the nature of the event. In the same way that our subject in the text mistook poison for wine, this subject misapprehends the physical nature of cavity-filling or childbirth labor. But it is also plausible to argue that the person understands the nature of the event and simply does not “properly” apprehend her reactions to it (or desires).

The ambiguity is caused of course by the fact that we can describe both people’s instrumental ends and desires at different levels of generality. Because it is conventional to think that we do not have a true taste for “the liquid in the glass” (rather than for wine or poison), we typically judge that misjudgment as narrowly instrumental. But if we recognize that the subject got (broadly) what he wanted (a good wine) but misestimated how much he would like it, we are more prone to say that he misapprehended his tastes rather than the good. But of course he was not (necessarily) fixated on wine itself. Wine may have been a (broad instrumental) vehicle to a certain sort of consumption happiness and he was simply misinformed that this good would give it.

Here is another way of apprehending the ambiguity: If we (try) to argue that one subject misunderstood something about the external world and one something about herself, we will inevitably be in some trouble (because we have too quickly assumed the “self” is a subject, not an object). It just is not clear whether the party “really” misunderstood her “reaction” to cavity-filling or
If, instead, though we say that any desire that turns out to make us unhappy is inadequately informed,\(^{111}\) in the sense that we did not have the information that it would make us unhappy,\(^{112}\) then the role of "information" is more

labor, or "really" misunderstood what these external stimuli do to the body's pain receptors.

Even if we think that it is best to picture a person who misestimates how painful she will find a procedure or event as misinformed about the event rather than about her preferences, there is still a thorny problem we need to figure out before we know what information a properly informed person should have. Generally, it appears that people overestimate the pain they will report, but underestimate the pain they end up reacting to behaviorally. Thus, dental patients interviewed before, immediately after and three months after a dental appointment overestimated reported pain, but women ask for anesthesia during labor far more often than they expect to, thus indicating that they expect to feel less pain than they actually do. Compare Gerry Kent, Memory of Dental Pain, 21 PAIN 187 (1985), and Arnoud Arntz, Mareleen vanEck & Monique Heijmans, Predictions of Dental Pain: The Fear of Any Expected Evil is Worse than the Evil Itself, 28 BEHAV. RES. & THERAPY 29 (1990), with Jay J.J. Christensen-Szalanski, Discount Functions and the Measurement of Patients' Values: Women's Decisions During Childbirth, 4 MED. DECISION MAKING 47 (1984). I am therefore not sure what the fully informed person actually knows about prospective pain estimation, even assuming that "mistaken" preferences about situations that cause physical pain involve simple mistakes in apprehension of the true nature of events external to the subject.

111. This appears to be the position that Kaplow and Shavell ultimately adopt, though they pay little attention to the question that I ultimately raise in the text: can we adopt this position without retreating from preference utilitarianism back to the sort of hedonic utilitarianism that preference utilitarianism discredited? See KAPLOW & SHAVELL, supra note 1, at 411-13.

112. Hedonic psychologists have obviously been most interested in these sorts of "errors"—ones that are conventionally most readily, if not wholly accurately, characterized as misapprehensions of one's own future internal psychological reactions rather than misapprehensions about the nature of external objects. They have classified a host of ways in which people mistakenly predict how they will react to a situation. Given that choices and preferences are formed on the basis of such predictions, the choices will systematically not work out as the subjects hope.

Some of the conventional psychological conclusions are interesting because of their generality. Most notably, subjects systematically predict that circumstantial changes (in their personal lives, in the social world and environment) will have a larger effect on their happiness than the precise same circumstantial shifts have had on their life when they have actually occurred in the past. See George Loewenstein & Shane Frederick, Predicting Reactions to Environmental Change, in ENVIRONMENT, ETHICS, AND BEHAVIOR 52 (Max H. Bazerman et al. eds.,1997).

More particular infirmities of judgment are also striking: people overestimate how afraid they will actually be in stressful situation. See S. Rachman, Panics and their Consequences, in PANIC: PSYCHOLOGICAL PERSPECTIVES 259 (S. Rachman & Jack D. Maser eds., 1988). They
ILLUSORY THAN REAL. CHOICES THAT MAKE US HAPPY ARE ONLY THOSE CHOICES THAT MAKE US HAPPY. IF THEY MAKE US UNHAPPY, IN FACT, WE WOULD NOT HAVE MADE THEM. TO THE EXTENT THAT WE THINK IT A GOOD IDEA TO SATISFY ONLY THOSE PROSPECTIVE DESIRES THAT ACTUALLY TURN OUT TO MAKE US HAPPY, THEN THERE IS NO INDEPENDENT ROLE TO SATISFYING PREFERENCES: WE SHOULD SIMPLY DO WHAT IN FACT MAKES US HAPPY. TRULY KNOWING WHAT MAKES US HAPPY, THOUGH, REQUIRES THAT WE HAVE THE ABILITY TO MEASURE HEDONIC UTILITY. WE CAN NO LONGER HOPE TO DETERMINE WHETHER WE WERE HAPPY BY REFERENCE TO WHETHER WE HAVE HAD DESIRES FULFILLED OR BEEN (IN SOME GLOBAL SENSE) “INFORMED.” BUT OF COURSE, IT MAKES PERFECTLY GOOD SENSE THAT THERE IS NO INDEPENDENT UTILITARIAN REASON TO SATISFY DESIRES. AS BRANDT NOTED, THERE IS REALLY NO INTRINSIC REASON TO CARE WHETHER ONE’S DESIRES ARE SATISFIED, EXCEPT INSOFAR AS THE FACT THAT ONE DESIRES SOMETHING SIGNALS SOME FURTHER FACT (MOST ESPECIALLY THAT ONE IS LIKELY TO LIKE EX POST WHAT ONE DESIRED EX ANTE). Thus, says Brandt, imagine that one said (and felt) at six, “I want to ride the roller coaster on my 50th birthday.” Not only does the fact that one had that desire not dictate what one does, it actually provides no independent reason to ride the roller coaster at all. Unless riding the roller coaster will make you happy, or help make the world more the way you now wish it would be, the mere fact that one would be satisfying a desire is wholly irrelevant.113

underestimate how vulnerable they will be to social pressures to conform if they place themselves in situations in which conformist pressures will exist. See Robert J. Wolosin, Steven J. Sherman & Annie Cann, Predictions of Own and Other’s Conformity, 43 J. Personality 357 (1975). Not surprisingly, they overestimate the degree to which they will be able to avoid giving in to addictive drug cravings or the urge to splurge at the shopping mall, so that initial choices to use drugs or to go shopping might prove hedonically troublesome. See George Loewenstein, A Visceral Account of Addiction, in Getting Hooked: Rationality and Addiction 235 (Jon Elster & Ole-Jorgen Skog eds., 1999) (on drugs); Stephen J. Hoch & George F. Loewenstein, Time-Inconsistent Preferences and Consumer Self-Control, 17 J.Consumer Res. 492 (1991) (on shopping).

113. See Brandt, supra note 3, at 249 (1979). This point is made quite forcefully in Thomas Scanlon, Value, Desire, and Quality of Life, in The Quality of Life 185,192-194 (Martha C. Nussbaum & Amartya Sen eds., 1993) (relying in part, but not wholly, on Brandt’s argument).

There are of course, many reasons that a person might want others to respect her preferences even if they are not hedonically satisfying and I will not detail them here. For now, I want to make a relevant preliminary point. It is really not inevitably the case that we must disclaim any choice that turns out (hedonically) badly. We might still both embrace the initial choice, and at some
meta-choice level, desire that all choices that were made under “adequate informational” conditions be met, even if we come to recognize that some other set of choices might have made us happier. But the decision to “ratify” hedonically imperfect choices cannot be grounded in the sense that doing so will maximize well-being so much as it maximizes the fulfillment of same sorts of autonomy ends.

Let us take a concrete case: Suppose, quite plausibly, that naïve (under-informed?) subjects do not adequately foresee the degree to which they might adapt to life circumstances that prospectively seem quite dismal (or that they do not anticipate that they will feel less elation once they have actually experienced great events than they expect to feel). Thus, for instance, assistant professors facing tenure decisions underestimate how happy they will be in five years if denied tenure and overestimate how happy they will be if granted it. See Daniel T. Gilbert et al., *Immune Neglect: A Source of Durability Bias in Affective Forecasting*, 75 J. PERSONALITY & SOC. PSYCHOL. 617 (1998). Prisoners expect to be much unhappier in solitary confinement than they prove to be. See Peter Suedfeld et al., *Reactions and Attributes of Prisoners in Solitary Confinement*, 9 CRIM. JUST. & BEHAV. 303 (1982). See also Elaine M. Sieff, Robyn M. Dawes & George Loewenstein, *Anticipated versus Actual Responses to HIV Test Results*, 112 AM. J. PSYCHOL. 297 (1999) (finding that people anticipate more distress given a positive HIV test than they actually experience and anticipate less distress given a negative test result than they actually experience). They might then make prospective choices (e.g., pull the plug if you can find a plug; do not get a test to determine HIV status) based in part on a simple lack of cognitive knowledge about the (supposed) prevalence of hedonic adaptation.

A subject might well find that learning information about hedonic adaptation shifted his choices, and even believe that the choices he would make if more informed about hedonic adaptation processes, would lead him to be happier, or at least to report higher overall levels of lifetime happiness, but still not believe the naïve choices are really worse. For it is possible of course, to respond to the information by choosing in a non-welfarist way. The subject may feel that the reality of adaptation serves, above all, to blunt the significance of choice in life generally (however things turn out, you will feel less good or bad about it than you expect). The “lesson” that those who recognize hedonic adaptation should then perhaps learn is to develop a more generalized indifference to choices. The subject who resists this “lesson” may thus wish to make all (particular) choices deluding himself into exaggerating their hedonic importance, rather than to undergo the less particularized losses that might be caused if he became more indifferent to his choices and more accepting of fate. Moreover, subjects may believe that it is worse “to be happy” because one’s expectations have been lowered—even if it really is a form of happiness—than to maintain higher expectations.

Plainly, one can—but need not—translate these decision rules into welfarist rules of the form (“seek a stronger commitment to the importance of choice only if that commitment itself is hedonically worthwhile” or “seek higher expectations only if having higher aspirations is hedonically satisfying”). Familiar commensurability problems loom large: it may be very difficult to compare the sort of happiness one gets from accepting one’s (more constrained) lot with the sort one gets from maintaining high expectations. For a good discussion of related issues, see Jon Elster, *Sour Grapes—Utilitarianism and
There is also no necessary connection between increasing access to information and hedonically favorable choice-alteration. Increasing information may make people more likely to regret their choices on some occasions and be irrelevant on others. If the goal of “correcting” desire is to increase the probability that people will be made happy by getting their desires satisfied, then it is an empirical, not a logical, claim that increasing the flow of accurate information will best correct desires (in the apt fashion). Small children are (mis)informed that Santa will not show up till they fall asleep; this misapprehension may lead them to go to sleep (and be less cranky on Christmas day).  

More generally, the relationship between being more globally informed and being happy is tenuous at best; but if all we want to ensure is that subjects are informed enough to change choices to ones that will make them happier, we do not really care about either increasing information, per se, or choice, per se, but only in happiness. The (controversial) literature on depressive realism suggests not only that information (generally) need not be pro-hedonic. It suggests as well that high levels of cognitive realism (well-processed information about the world) at least correlates with and perhaps (at least to some degree) causes depression. People suffering from depression seem (in certain older experiments) to be less self-deluded than non-depressives about the low levels of control they actually exercise over chance events, like winning prizes in what are

the Genesis of Wants, in UTILITARIANISM AND BEYOND, supra note 45, at 219, 231-38.

114. The example comes from Peter J. Hammond, Utilitarianism, Uncertainty and Information, in UTILITARIANISM AND BEYOND, supra note 45, at 85, 101. Hammond notes that it might nonetheless be appropriate (in terms of “respecting autonomy”) that (at least certain) decision makers be more fully informed, whether or not this increases ex post utility.

115. For a parallel critique and example, see ALLAN GIBBARD, WISE CHOICES, APT FEELINGS 20-21 (1990). For his fuller discussion of the ambiguities and infirmities of full information idealized preference satisfaction theories, see id. at 18-22, 183-88.

116. Most of the studies, to the degree they are persuasive at all, speak to the correlation issue, not the causation issue.
essentially lotteries—thus, the effect of increasing the level of (effectively processed) information that people have about their lack of control in these quasi-lotteries is at best ambiguously related to happiness. Similarly, classical experiments seemed to indicate that non-depressives (unduly) discount negative feedback about their traits or performance and simultaneously overestimate the veracity of positive description while depressives absorb negative and positive information in a less biased fashion. Once more, it is by no means obvious that clarifying to subjects that the people who see them, warts and all, are not as petty or myopic as they might first imagine will clearly improve their subjective state.\textsuperscript{117}

Perhaps even more telling, it is not clear generally whether it is hedonically wise to develop more sophisticated and discerning tastes.\textsuperscript{118} Once one “learns” to differentiate very good wines from poorer ones, the poorer ones taste disappointing, and one may not be happier as a result. Moving to a meta-choice preference-utilitarianism seems unavailing: What does it mean to choose to be the sophisticated person or the unsophisticated one? Which

\textsuperscript{117} For a typical study purporting to document the existence of such “depressive realism,” see M.M. Moretti et al., \textit{Self-Referent Versus Other Referent Information Processing in Dysphoric, Clinically Depressed, and Remitted Depressed Subjects}, 22 \textit{PERSONALITY \& SOC. PSYCHOL. BULL.} 68 (1996) (finding that nondysphoric subjects provided with both negative and positive social reactions directed at them discounted how informative the negative reactions were, while dysphoric subjects rated positive and negative reactions as equally informative). Classic works on “depressive realism” include Shelley E. Taylor & Jonathon D. Brown, \textit{Illusion and Well-Being: A Social Psychological Perspective on Mental Health}, 103 \textit{PSYCHOL. BULL.} 198 (1988); Walter Mischel, \textit{On the Interface of Cognition and Personality: Beyond the Person-situation Debate}, 34 \textit{AM. PSYCHOL.} 740 (1979); Lauren B. Alloy & Lyn Y. Abramson, \textit{Judgment of Contingency in Depressed and Nondepressed Students: Sadder but Wiser?}, 108 \textit{J. EXPERIMENTAL PSYCHOL.: GEN.} 441 (1979); and Lauren B. Alloy & Lyn Y. Abramson, \textit{Learned Helplessness, Depression and the Illusion of Control}, 42 \textit{J. PERSONALITY \& SOC. PSYCHOL.} 1114 (1982). For a good, if slightly dated, review essay, see Lauren B. Alloy & Lyn Y. Abramson, \textit{Depressive Realism: Four Theoretical Perspectives, in COGNITIVE PROCESSES IN DEPRESSION} 223 (Lauren B. Alloy ed., 1988). One of the better pieces criticizing the “depressive realism” hypothesis is C. Randall Colvin, Jack Block & David C. Funder, \textit{Overly Positive Self-Evaluation and Personality: Negative Implications for Mental Health}, 68 \textit{J. PERSONALITY \& SOC. PSYCHOL.} 1152 (1995).

\textsuperscript{118} For a good account of this (and other) problems with full information accounts of well-being, see David Sobel, \textit{Full Information Accounts of Well-Being}, 104 \textit{ETHICS} 784 (1994).
"self" gets to make the choice? But, if instead, we must measure whether the person was happier before or after she became a sophisticated wine snob, we need to return to hedonic utilitarianism.

It is equally unclear how we might describe preferences as more prudently formed without taking a peek at the answer to the question that preference utilitarianism seeks to duck: is the subject actually happy in the state he has arrived at? It is hard to establish independent "prudence" criteria that assure us that the subject has chosen prudently if, despite some version or other of full information both about the nature of the object states he must evaluate and the nature of his ultimate reactions to those states, he nonetheless prefers something that we believe is dysphoric. One of the most consistent lessons of "hedonic psychology" is that people make different choices in distinct psychological background states. We could judge one choice, rather than another, as more prudent only if we knew it were associated with a better outcome. (And, of course, for the utilitarian, "better" could only mean hedonically superior).

Some of the choice-influencing psychological contexts are fairly predictable. For instance, when in a "cold" (or non-aroused, dispassionate) state, we tend to mis-predict the choices we will make in a "hot" state (under duress or temptation). People overestimate the strength of their willpower in hot situations — so that a choice to vacation

---

119. One of the clearest discussions of this familiar point is in Frank Hahn, On Some Difficulties of the Utilitarian Economist, in UTILITARIANISM AND BEYOND, supra note 45, at 187, 190-95. The same point is raised in Kelman, supra note 108, at 786-87.

120. For a good summary, see George Loewenstein & David Schkade, Wouldn't It Be Nice? Predicting Future Feelings, in WELL-BEING, FOUNDATIONS, supra note 5, at 85.

121. See id. at 93. Similarly, smokers radically underestimate how likely it is that they will smoke in five years. See Barbara S. Lynch & Richard J. Bonnie, Towards a Youth-Centered Prevention Policy, in GROWING UP TOBACCO-FREE: PREVENTING NICOTINE ADDICTION IN CHILDREN AND YOUTHS 3 (1994). Because we are fairly certain that the later "hot" (addicted?) decision to smoke is a literally self-destructive one that would both typically be disclaimed and regretted and is difficult to defend hedonically, we readily assume that the earlier decision to smoke is imprudent. Similarly, people (counter-hedonically) choose to have unsafe sex in the "heat of the moment" even though they planned, before foreplay, to practice safe sex. See Ron S. Gold, On the Need to Mind the Gap: On-Line versus Off-Line Cognitions Underlying Sexual Risk-Taking, in THE THEORY OF REASONED ACTION: ITS APPLICATION TO AIDS-PREVENTIVE BEHAVIOR
in Las Vegas made by the person who expects to gamble moderately might be thought imprudent, just as the decision to gamble extensively while there might be thought to render choices to stay home on other occasions imprudent. Similarly, “normal-weight” shoppers buy more food when they enter the store hungry than not, but it is not clear why—on purely procedural grounds, divorced from reference to the hedonic outcomes of the distinct choices—we would say that buying more when hungry (rather than less when satiated) is imprudent. Thus, we might smuggle in some substantive hypothesis about how happy the subject would or should be living with her particular decisions (Americans eat too much for their own good, Americans buy too much for their own good). After all, one could readily argue (a priori, focused more exclusively on procedure) that one will eat when hungry so that one’s purchases in a satiated state more poorly reflect the desires one will have at the relevant point of consumption.

The abstract notion that the “cold” choice is the more prudent one may simply reflect a (more or less well-analyzed) belief that hot choices work out more poorly, hedonically. It is not clear, conceptually, what else it could reflect. It is almost surely the case that the “intuitive” purportedly procedural preference for “cold” choices is entirely substantive. It may be (what I have been calling) “strongly” perfectionist (e.g., grounded in the belief that certain choices are more likely to advance our “real” interests than others). As a strong perfectionist intuitive theory, it is substantively Puritanical: the cold chooser eats less and is more sexually inhibited. Those are good things only given a strong theory of human flourishing. It may also

227 (Deborah J. Terry et al. eds., 1993). But think about an area where there might be high—but far less absolute—consensus that the “hot” choice is the bad one: subjects may well underestimate how likely it is that they will engage in sex once they get into arousing situations, like dates. See Loewenstein & Schkade, supra note 120, at 93. But unless one is sure that the decision to have sex during the aroused period is hedonically worse (or worse in some strong perfectionist sense if one is willing to disclaim subjective utilitarianism more thoroughly) than the decision to refrain in the cold period, it is not clear why one would call one decision more prudent than the other.

122. Similarly, they typically underestimate the level of courage they might show in the face of emergencies. See Loewenstein & Schkade, supra note 120, at 93. Thus, their choices to avoid certain high-meaning, high-risk settings (e.g., going off to fight wars one deeply believes in) might also be imprudent in the relevant sense.
be what I have been calling "weakly perfectionist:" certain reflective or evaluative positions are thought to be better than others. The aspect of our observing selves that looks back from a distance (in reflective equilibrium) or forward (from a detached, uninvolved planner's perspective) is thought to be the part whose assessments count, not the self that is (arguably) most engaged and present in the moment.

I suspect most people intuitively understand that we form distinct preferences when aroused and when calm. Less well understood is that other distinct psychological states have the same sorts of influence on choice and preference formation. Moreover, I suspect that reactions to the prudence of making choices under one or the other of these less transparent influences are considerably less strongly held or widely shared than (what I see as inadequately self-critical but widely shared) reactions to the choice between "cold" and "hot" decision backdrops. For instance, people choosing a large number of goods at a particular point in time exhibit what might (pejoratively) be called a diversification bias, compared to those who choose from among the same goods seriatim. Thus, if students are asked to pick what snack they will eat in each class over the next three weeks, they are more likely to choose different snacks for each week's class than if they are asked to choose the snack for each day at the beginning of that day's class.123

123. More generally, it is possible, but I think unavailing, to identify prudent choices with those that remain most stable over time, once alternatives have been reflected upon. Certainly, moral philosophers like Rawls have argued that it is better to trust moral sentiments that are stable in reflective equilibrium than mere intuitions. See John Rawls, A Theory of Justice 48-51 (1971). I think, though, that stable judgments may well be more hedonically problematic than unstable ones. Think, once more, about the food disgust examples: once we know the provenance of food we think of as disgusting, we might well be unable to eat it even though it would be in our hedonic interest to do so (e.g., if we would otherwise starve on some hypothetical desert island). Not only is it bad to be more informed (of the food's provenance), it is bad to arrive at the stable equilibrium point. For a fuller discussion, see Mark Kelman, Problematic Perhaps, but Not Irrational, 54 Stan. L. Rev. 1273, 1284-86 (2002).

To the degree that people regret, *ex post*, the decision to diversity, and to the extent we believe that this regret bespeaks actual hedonic loss, then the seriatim choice could be described as more prudent. But my point is that it could only be described as more prudent in the very limited and ultimately non-procedural sense that it appears to have turned out to be hedonically superior. I cannot imagine that any of us have an instinct that seriatim choices are more deliberate or deeply considered. To the contrary, if people share any weak intuition on this issue, I suspect it would be that the person who "plans" over a longer period is being more thoughtful and prudent than the person who makes one decision at a time.

Similarly, most of us are unaware of the degree to which curiosity drives our choices. Subjects are more likely to choose to receive the answers to ten geography questions rather than an attractive candy bar once they have seen and tried to answer all the questions than when they have seen only a single randomly sampled question (which might give them some idea about whether they already know the answers to the questions). They also underestimate how likely they are to choose to satiate their curiosity rather than appetite.\textsuperscript{125} Thus, choices made when one is not curious might differ from those made when curious, but it is difficult even to imagine why one set of choices would be viewed as more prudent than the other. If one tried to remain committed to procedural accounts, it is easy enough to imagine making the argument that curiosity is just the sort of short-term impulsive itch-that-needs-scratching that leads to imprudent decisions or just the opposite—that people who are curious are intellectually engaged and alive in just the way that a prudent decision maker should be.

In sum, we neither know what information we need nor when we are prudent without knowing whether the choices we have made—given the information we had, given the degree of prudence we showed—"worked out" for the best. Preference-utilitarianism cannot be rescued by invoking pure procedure.

---

\textsuperscript{125} See George Loewenstein, *The Psychology of Curiosity: A Review and Reinterpretation*, 116 PSYCHOL. BULL. 75 (1994); Loewenstein & Schkade, supra note 120, at 93.
Subjective welfarism is attractive both because it acknowledges what seems to be a relatively uncontroversial factual observation—people's tastes are indeed diverse—and a reasonable normative premise—individuals are ultimately subjects, legitimately concerned above all with living a life that is satisfying in their own eyes, rather than objects meant to meet some outside observer's goals. But we inevitably define philosophical positions—including subjective welfarism—in opposition to alternative views, and when we attempt to draw a sharp contrast between welfarist and perfectionist principles, I believe we fail.

This is true in part simply because the observer's decision to attend to welfare must ultimately be grounded in the perfectionist belief that it is worthwhile to advance welfarist goals, rather than other goals. But what I have emphasized in this paper is that it is more strikingly true because the concept of subjective welfare is at core jumbled and internally contradictory; only by recourse to what I have called "weak perfectionism" can it be made coherent.

Hedonic utilitarians—including the new hedonic psychologists in their most expansive, intellectually "imperialistic" moments—assume, quite unpersuasively, that people inexorably seek momentary pleasure (or more instances of "approaching" rather than "avoiding" their circumstances) as the end of life. We can reject this claim on both philosophical and empirical psychological grounds.

Philosophically, there are two core problems with the claim. First, it is incompatible with the subjectivism and respect for diversity that best justifies subjective welfarism altogether. People may reasonably seek to live lives that are neither dominated by pleasure nor approach. Second, it rests on a partially dissolved view of the person that is both unpersuasive and, more striking, incompatible with attending to any individual's welfare rather than the sum total of painful and pleasurable experiences in the world, regardless of their locus. People may well seek lives that are viewed as satisfying from a variety of more "integrated" perspectives. When we attempt to maximize an agent's welfare, what we can do (at best) is to maximize how well off she is from one such perspective, but our decision to attend to one of the numerous perspectives the agent could adopt reflects weak perfectionist ideals. Psychologically, the
claim is implausible given the wide variety of incommensurable and non-scaled positive and negative emotional states that we observe and the wide variety of biological roles that happiness and unhappiness play.

Preference utilitarians can avoid this difficulty in hedonic utilitarianism by remaining agnostic about whether people do or should seek any particular sensations. But satisfying preferences does not, by itself, make people well-off. An agent's preferences can be satisfied without affecting the agent at all (e.g., if she had died before they were satisfied), and, more commonly, satisfying them can fail to improve her experience because they are simply mistaken predictions of the hedonic impact of the chosen end-state. Preference utilitarians purport to solve this problem by respecting only procedurally valorized preferences—those preferences dubbed informed and prudent. But the psychology literature helps explicate a familiar philosophical point. Information is neither intrinsically valuable (given a subjective welfarist worldview) nor does getting more information inevitably make us choose things that make us happier. Moreover, although choices are distinct given distinct background conditions, it is not sensible to call the decisions made given one particular backdrop more prudent without regard to whether the hedonic impact of the decision is superior.

Thus, the decision to respect a certain class of preferences is either based on a covert perfectionist lionization of a certain procedural posture—people are more "fully realized," whether they are "happier" or not, if they make "cold" choices rather than "hot" ones, it fulfills our nature as highly "cognitive" creatures if we are as informed as we can be in developing desires—or more plausibly, based on peeking at the hedonic impact of the choice. A decision is prudent not because it was made in certain settings nor adequately informed because the decision maker had a certain level of information about the world. It is prudent or informed if it turned out well, by the lights of the hedonic utilitarian (whom the preference utilitarian rightly dismissed as being too covertly perfectionist to satisfy a true subjectivist's aims).

What that means is that hedonic utilitarianism—with its inevitable perfectionism—is ultimately where we must wind up if we are to remain committed to welfarism. The "observer" who purports merely to sit back, uninvolved, and
tote up subjective reactions will inexorably ask some questions and omit others, or "count" the answers to some questions as more expressive of the "sort of welfare that matters." Subjective welfare, the hero of utilitarianism and policy science, is a unicorn.