

Yale Law School Student Prize Papers

Student Prize Papers

Yale Law School

Year 2008

NON-ILLUSORY LOSSES: WHY PAIN
AND SUFFERING DAMAGES
SHOULD (JUST) BE ABOUT PAIN
AND SUFFERING

Lucy Wang
Yale Law School, lucy.wang@aya.yale.edu

**NON-ILLUSORY LOSSES:
WHY PAIN AND SUFFERING DAMAGES SHOULD (JUST) BE ABOUT PAIN AND SUFFERING**

**LUCY WANG
PROFESSOR CHRISTINE JOLLS
SPRING 2008**

Introduction

Recent psychological research suggests that people underestimate their ability to adapt to radical changes (“hedonic adaptation”). Consequently, we overestimate the impact that adverse events will have on our long-term happiness. This tendency to mis-predict future (un)happiness raises the legal question: Are juries overcompensating for pain and suffering of tort victims? Cass Sunstein¹ and Peter Ubel and George Lowenstein² argue that juries would be overcompensating according to purely hedonic conceptions of pain and suffering. Nonetheless, victims deserve compensation for their non-hedonic losses.

The controversy over pain and suffering compensation is an instance of the broader conflict between three general theories of happiness: hedonic wellbeing, desire fulfillment, and objective list. Sunstein and Ubel and Lowenstein [hereinafter “U&L”] argue that hedonic adaptation vitiates victims’ claims to lost hedonic wellbeing. Nonetheless, we still have a strong moral intuition that such injuries are worthy of compensation. Therefore, hedonic wellbeing is insufficient justification for tort compensation.

With this problem in mind, Sunstein and U&L search for alternative theories of compensation. U&L pick the desire satisfaction account. According to this theory, the fact that victims still prefer their pre-injury condition justifies compensation. Victims’ preferences are valuable independent of whether those preferences produce hedonic wellbeing. Meanwhile, Sunstein settles on the objective list theory. According to this theory, some states of being are intrinsically valuable, even if nobody prefers them. What makes these states valuable is an independent, universal normative theory.

¹ Cass R. Sunstein, *Illusory Losses*, Draft 07/09/07.

² Peter A. Ubel & George Loewenstein, *Pain and Suffering: It’s Not (Just) About Pain and Suffering*; *see also* George Loewenstein & Peter A. Ubel, *Hedonic Adaptation and the Role of Decision and Experience Utility in Public Policy*, Conference on Happiness, and Public Economics (London, Sept. 22 – 23, 2006) (on file with author).

In this paper, I will argue that the Sunstein and U&L argument is flawed at both stages. First, hedonic adaptation does not vitiate claims of hedonic loss to the degree that authors imply. Catastrophic losses are worthy of compensation, and hedonic wellbeing captures the right theory of compensation. Second, neither alternative theory provides a satisfying justification for tort compensation. While desire satisfaction and objective list theories may be appealing in other contexts, they are inappropriate as applied to tort compensation. This paper will begin with an introduction to the three theories of happiness. Thereafter, the paper will examine the alleged problem of hedonic adaptation and the proposed solutions.

Three Accounts of Happiness

There are three broad conceptual approaches to understanding happiness, or to use a less loaded term, individual wellbeing. Derek Parfit identifies these as: hedonic wellbeing, desire fulfillment, and objective list.³ Each account answers the question “what makes someone’s life go best[?]”⁴ This Part will define each theory and then clarify two basic confusions about their differences: the role of pleasure and level of paternalism.

i. Definitions

First, the hedonic wellbeing account posits that the best life is one filled with the best subjective mental states.⁵ According to this theory, only facts about an individual’s mental state matter, not facts about the external world. Hedonic accounts vary, however, in the choice of metric for what constitutes the best mental state. Parfit identifies two potential metrics: “narrow hedonism” and “preference hedonism.”

³ DEREK PARFIT, REASONS AND PERSONS app. I at 493 (1984). I will refer to the first account as “hedonic” rather than “hedonistic.”

⁴ *Id.*

⁵ I will follow Paul Dolan and Mathew P. White in describing the aim of hedonistic accounts in terms of “subjective mental states” rather than “happiness.” Paul Dolan & Mathew P. White, *How Can Measures of Subjective Well-Being BE Use to Inform Public Policy?*, 2 PERSPECTIVES PSYCHOL. SCI. 71 (2007). This paper will treat all three accounts as theories of happiness.

According to “narrow hedonism,” mental states vary along a single axis from pleasure to pain.⁶ The best states are those that produce the most pleasure. Martha Nussbaum attributes this view to Jeremy Bentham: “For him, pleasure simply must be a single homogeneous sensation, containing no qualitative differences.”⁷ Nussbaum and Parfit both reject this narrow view as overly simplistic.⁸ As Parfit explains, the diversity of human experience cannot be analyzed in terms of a single component: “Compare the pleasures of satisfying an intense thirst or lust, listening to music, solving an intellectual problem, reading a tragedy, and knowing that one’s child is happy. These various experiences do not contain any distinctive common quality.” To be psychologically plausible, a hedonistic account must acknowledge the multi-dimensionality of subjective experience. While pleasure is an important component of subjective experience, a state with more pleasure is not necessarily better than one with less pleasure.

Consequently, a more nuanced hedonic account acknowledges that individuals have complex preferences regarding their own mental states. According to “preference hedonism,” the best mental states are simply those that individuals desire the most.⁹ Preference hedonism takes such desires as primitive; the theory does not dictate what states people should desire. Within preference hedonism, an individual may desire a mental state even though it produces more pain than pleasure. Nussbaum, lists, for example many plausibly desirable but painful states: “the grief that expresses love, ... the compassion that shares the pain of the suffering person, the anger that says, ‘This is deeply wrong and I will try to right it.’”¹⁰ Unlike narrow hedonism, preference hedonism does not deny the rationality of preferring a painful state over a pleasurable

⁶ PARFIT, *supra* note 3 at 493.

⁷ Martha C. Nussbaum, Who is the Happy Warrior?, Philosophy Poses Questions to Psychology, Conference at University of Chicago (June 1-2, 2007) at 2.

⁸ PARFIT, *supra* note 3 at 493.

⁹ PARFIT, *supra* note 3 at 493.

¹⁰ Nussbaum, *supra* note 7 at 16.

one. Thus, if Nussbaum's "happy warrior" prefers the warrior mental state above all alternatives, he is living the best hedonist life in battle. Parfit cites a similar example in Sigmund Freud's real end-of-life choices: "Near the end of his life Freud refused pain-killing drugs, preferring to think in torment than to be confusedly euphoric. Of these two mental states, euphoria is more pleasant. But on Preference-Hedonism thinking in torment was, for Freud, a better mental state."¹¹ Given Freud's preferences, torment was the rational choice under preference hedonism. Of course, for someone with different preferences, euphoria would have been the rational choice. In this way, preference hedonism defers to individual preferences in judging mental states.

For individuals to have well formed preferences, however, mental states must be commensurable at some level. Preference hedonism, therefore, denies Nussbaum's more radical claim that life's experiences "cannot be rendered commensurable on *any* quantitative scale."¹² It is beyond the scope of this paper to defend the truth of the fundamental hedonic hypothesis. At this point, I only want to assert that commensurability is at least intuitively plausible. As Paul Dolan and Mathew P. White describe, comparing subjective states is a part of the everyday deliberative process: "[W]hen individuals are asked why they want to achieve various objective goals, such as losing weight or saving money, they often conclude that it is because achieving these goals will improve their SWB [subjective wellbeing]."¹³ If subjective experiences were truly incommensurable, the idea of "improving" one's subjective wellbeing would be incoherent. Thus, the fact that people do employ such practical reasoning is *prima facie* evidence that preference hedonism is psychologically plausible. In this paper, I will focus on preference hedonism rather than narrow hedonism.

¹¹ PARFIT, *supra* note 3 at 494.

¹² Nussbaum, *supra* note 7 at 16 (emphasis added).

¹³ Dolan & White, *supra* note 5 at 72.

Preference hedonism is closely related to the second broad theory of individual wellbeing: the desire-fulfillment account.¹⁴ According to this theory, the best life is one that fulfills an individual's greatest desires, where the scope of relevant desires includes "all preferences about our own lives."¹⁵ Of course, many of our desires are to achieve certain subjective mental states, *e.g.*, "I want to be in love." In those cases, preference hedonism and desire fulfillment converge. In other cases, however, our desires are about objective states of the world, *e.g.*, "I want my spouse to be faithful." In those cases, the two accounts can diverge if one is deceived about the world. For preference hedonism, all that matters is that I achieve the desired mental state, *e.g.*, I am satisfied in my marriage because I mistakenly believe that my spouse is faithful. For desire-fulfillment, on the other hand, "It is bad for me if [my] desire is not fulfilled, even if I falsely believe that it is."¹⁶ Thus, the desire-fulfillment theory defers to internal preferences to define the conditions of the best life. But the theory refers to external facts to verify whether those conditions have been met.

Finally, there is the objective list account. According to objective list theories, "certain things are good or bad for us, whether or not we want to have the good things, or to avoid the bad things."¹⁷ The best life is one that achieves the good and avoids the bad. Unlike preference hedonism and desire fulfillment, the objective list account is completely external. Instead of deferring to individual preferences, the account relies on an independent theory of the good; the account "appeals directly to what it claims to be facts about value."¹⁸ Consequently, the objective list account imposes a uniform metric of the good life upon all individuals. In this paper, I will

¹⁴ PARFIT, *supra* note 3 at 493

¹⁵ *Id.*

¹⁶ *Id.* at 494.

¹⁷ *Id.* at 493.

¹⁸ *Id.* at 499.

focus on two objective list theories: the Aristotelian theory of happiness as eudaimonia and its modern successor, the Capabilities Approach.

ii. Role of Pleasure

The three theories of happiness differ in their choice of reference point. Preference hedonism measures an individual's life against his own desires about his subjective wellbeing. Desire fulfillment defers to individual desires about the world more broadly. Objective list utilizes independent criteria of the good life. What does not distinguish the theories, however, is the role they assign to pleasure.

Specifically, it is a misnomer that hedonic accounts prioritize pleasure while objective list accounts discount it. Neither theory mandates a commitment for or against pleasure. As discussed above, one hedonic account, narrow hedonism, measures subjective mental states by the amount of pleasure they produce. One could easily, however, engineer an opposite metric, whereby mental states are valued by the amount of pain they produce.¹⁹ Although this would be an unappealing theory of happiness, it is still a hedonic account because it only takes into consideration only facts about subjective wellbeing. A more plausible hedonic account is preference hedonism, in which pleasure plays no set role at all. Within preference hedonism, pleasure matters only to the extent that individuals prefer it. (The same would be true of desire fulfillment accounts.) Thus, while pleasure is a subjective mental state, hedonic accounts need not be defined by pleasure; pleasure can play a range of roles or no fixed role at all.

Similarly, objective list theories need not assign any intrinsic value to pleasure. Both Aristotelian ethics and the Capabilities approach, however, emphasize the importance of pleasure to human life.²⁰ The *Nicomachean Ethics* famously presented two competing views of pleasure:

¹⁹ Assuming for the moment that pain and pleasure are two ends of one spectrum.

²⁰ See, e.g., *Nicomachean Ethics*, 1099a7-20.

(1) that the highest good is a form of pleasure and (2) that pleasure is not itself a good, but rather something that accompanies virtuous activity.²¹ Under either interpretation, pleasure has a close relationship to Aristotle's ideal of eudemonia. Likewise, the Capabilities Approach explicitly lists as the third human functional capability: "Being able to avoid unnecessary and nonbeneficial pain and to have pleasurable experience."²² Therefore, for both hedonic wellbeing and objective list theories, pleasure can play a central role, although not necessary the defining role. The difference is not how much weight is given to pleasure, but rather the source of its value. Preference hedonism (and desire fulfillment) value pleasure to the extent that individuals desire it. Objective list, on the other hand, places a fixed value on pleasure according to its contribution to a fulfilling human life.

iii. Level of Paternalism

Another point of confusion stems from the level of paternalism that each theory entails. As Dolan and White argue, the primary objection to objective list theories is paternalism.²³ Mark Kelman, however, counters that hedonic accounts are just as paternalistic.²⁴ There is truth on both sides; objective list and hedonic wellbeing are both paternalistic, but on different levels. Objective list theory is paternalistic about the ends of human life, whereas hedonic wellbeing (and analogously desire fulfillment) is paternalistic about only about the means.

Objective list theories provide a normative account all humans should value in their lives. Nussbaum describes the basic project: "We are not pretending to discover some value-neutral

²¹ The first position is developed in Book VII, whereas the second position is developed in Book X of the *Nicomachean Ethics*. See generally, Richard Kraut, *Aristotle's Ethics*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2007), <http://plato.stanford.edu/entries/aristotle-ethics/> (discussing Aristotle's two treatments of pleasure).

²² Martha C. Nussbaum, *Human Functioning and Social Justice: In Defense of Aristotelian Essentialism*, 20 POL. THEORY 202, 222 (1992)

²³ Dolan & White, *supra* note 5, at 74.

²⁴ Mark Kelman, *Hedonic Psychology and the Ambiguities of "Welfare,"* 33 PHIL. & PUB. AFFAIRS 391 (2005). For a similar argument, see Lowenstien & Ubel, *supra* note 2, manuscript at 1.

facts about ourselves, independent of all evaluation; instead we are conducting an especially probing sort of evaluative inquiry.”²⁵ The objective list theorist takes a normative stances on valuable human goals and actions, and even human emotions. Nussbaum explains: “I think emotions are positive or negative, in the sense relevant to normative thinking, according to the correctness of the appraisal or evaluations they contain.”²⁶ An individual who desires different goals or different subjective experiences is simply wrong about what constitutes human happiness.

In contrast, hedonic wellbeing defers to individual preferences in defining the conditions of the best life. Individuals cannot be wrong about which subjective mental states constitute the best life. Yet people can be wrong about whether certain choices will get them to the desired state. That is, people often make choices based on inaccurate predictions of their own future subjective responses. Kelman argues that if hedonic theorists want to label such choices “imprudent,” they must first “peek ahead at the answer” to see whether the more prudent choice actually makes people happier.²⁷ In doing so, hedonic theorists must make normative assessments of the two choices. Therefore, Kelman argues, hedonic theorists cannot escape normative judgments of the same sort that objective list theorists make.

Kelman is right that hedonic theorists are making a normative judgment when they say that one choice is more “prudent.” Hedonic theorists can criticize the rationality of “imprudent” choices, however, without judging between ultimate ends. The hedonic theorist is not saying that individuals *should* prefer the one subjective mental state over another (because one is intrinsically better). That kind of judgment would be paternalistic in the same way that objective list theories are paternalistic. Instead, the hedonic theorists is saying that people on average *do*

²⁵ Nussbaum, *supra* note 22, at 214.

²⁶ Nussbaum, *supra* note 7, at 14.

²⁷ Kelman, *supra* note 24, at 396-97.

prefer one over the other. People who make the “imprudent” choices fall in one of two camps. First, some people may have idiosyncratic preferences. They may actually prefer the subjective state produced by the “imprudent” choice. The “imprudent” choice is for them prudent. These people, however, must be in the minority (or else the ordering of average reported satisfaction would be reversed). Therefore, a majority of people would actually prefer the subjective state produced by the “prudent” choice. But they make the wrong choice because they miscalculate their future subjective wellbeing. In this way, hedonic wellbeing is paternalistic about the means of achieving subjective wellbeing without being paternalist about what constitutes happiness for any individual.²⁸

In essence, the hedonic theorist is like an investment adviser who tells you not to buy a stock even though you believe it will be a winner. The difference, however, is that lay investors do not feel that they have any special ability to predict future stock values. Consequently, lay investors often seek and rely on expert predictions. On the other hand, people do feel that they have special ability to predict their own future subjective states. So when experts tell them the likely result of their choices, people bristle at the intrusiveness of the advice. Kelman is accurately capturing a sense in which people feel hedonic accounts are paternalistic. But, the argument is misleading because hedonic accounts offer prescriptive advice only about means not ends; this level of paternalism is present among professionals in any field of predictive science.

Tort Victims Have Significant Claims for Hedonic Loss

Turning now to pain and suffering damages, Sunstein and U&L fear that hedonic wellbeing cannot fully justify the level of tort compensation that we believe victims morally deserve. Specifically, they worry that if victims hedonically adapt, they no longer have claims

²⁸ Analogously, desire satisfaction theorists can criticize people for making poor choices that will not actually produce the changes in the world that they predict they will.

for hedonic loss. While tort victims do not suffer permanently lower hedonic wellbeing, the process of hedonic adaptation is not costless. Consequently, tort victims do suffer serious short-term hedonic losses. Authors are right that jurors overestimate the duration of this loss. But jurors are likely duration insensitive in their willingness to compensate. Therefore, a correction for hedonic adaptation may not necessarily decrease tort compensation. Sunstein and U&L overstate the threat that hedonic adaptation poses to hedonic theories of compensation.

i. Hedonic adaptation is costly

Sunstein and U&L argue that hedonic adaptation eliminates any significant claims for hedonic loss that tort victims may have had. The thrust of the argument is this: if victims adapt hedonically, they suffer no hedonic loss. This argument falsely assumes that adaptation is costless. Before getting to the issue of whether juries overcompensate, it is important to clarify what hedonic losses victims of catastrophic injury do actually suffer. While victims may eventually recover to pre-injury levels of hedonic wellbeing, the short-term costs of hedonic adaptation are significant and worthy of compensation.

Sunstein provocatively entitles his piece, *Illusory Losses*. The title itself implies that victims' losses are non-existent rather than merely transient. To his credit, Sunstein initially adds the caveat that the "illusory" applies only to the duration of the loss: "losses...turn out to be illusory (*at least in the long term*)..."²⁹ Similarly: "[M]any apparently serious losses inflict relatively little in the way of *long-term* hedonic harm."³⁰ Thereafter, however, Sunstein's rhetoric gets muddy.

²⁹ Sunstein, *supra* note 1, at 8 (emphasis added).

³⁰ *Id.* at 9 (emphasis added).

The main problem is that Sunstein conflates short-duration with low-intensity. In the following hypothetical, for example, Sunstein casually throws in the extra assumption that the victim's hedonic losses are "very small" in magnitude:

Suppose, for example, that a plaintiff has lost two fingers or an arm, and the jury is asked to monetize the loss, including the pain and suffering associated with it. Because of the power of the psychological immune system, it is not implausible to think that the loss is *short-term and very small*.³¹

Sunstein does not offer evidence that losing one's arm is a relatively pain-free process. Given the trauma of amputation and the prevalence of phantom limb pain,³² this assumption seems questionable, if not implausible. While the "power of the psychological immune system" allows victims to recover relatively quickly, it does not prevent the immediate hedonic losses from occurring. Without more evidence, it seems a stretch to characterize the pain of amputation as "very small."

The leap from short-duration to low-intensity is ungrounded in hedonic adaptation research as Sunstein himself later acknowledges: "It is both true and important that even if long-term harms are not likely, the short-term harms might be severe....The only point is that juries are likely to exaggerate the long-term effects and to that extent award excessive damage awards."³³ So, in a more careful moment in the paper, Sunstein makes clear that the "only point" at issue is the duration not the magnitude of pain.

Unfortunately, Sunstein's carefulness is soon forgotten. In his climactic intuition-pumping hypothetical, he again implies that hedonic adaptation is costless:

³¹ *Id.* at 13 (emphasis added).

³² Almost all patients experience phantom limb sensation following amputation. A large majority of upper limb amputees also experience phantom limb pain, which can be excruciating and is often difficult to treat. *See, e.g.,* Carolien M. Kooijman et al, *Phantom Pain and Phantom Sensations in Upper Limb Amputees: An Epidemiological Study*, 87 PAIN 33 (2000) (reporting that 82% of upper limb amputees experience phantom limb pain).

³³ Sunstein, *supra* note 1, at 13.

Suppose that Jones loses the use of a leg; suppose too that the loss does not affect that [sic] Jones' self-reported happiness. *After a difficult but short period of adjustment*, Jones is as happy as he was before the loss. Suppose too that the effort to measure Jones' moment-to-moment happiness finds that he is no less happy than he was before. In other words, *Jones has experienced no hedonic loss*. Should the legal system therefore disregard Jones' injury?³⁴

The logic seems to be that if Jones's before and (long-term) after measurements of happiness match, he has not lost anything in the interim. Following that logic, one could compare the unemployment rates of 1999 and 2007 and conclude that the recession of 2001 did not cause any significant job losses.³⁵ On the other hand, Sunstein is not completely ignoring the adaptation period. He explicitly stipulates that Jones suffers through a "difficult" adjustment. But, if the adjustment is so difficult, Jones could not have "experienced no hedonic loss." Perhaps what Sunstein means is that Jones experienced no *significant* hedonic loss. Even for this modified conclusion, it must be the case that adaptation, while "difficult," is essentially costless. This is the unstated and unsubstantiated premise that underlies Sunstein's ultimate argument.

Thus, Sunstein uses hedonic adaptation research in a misleading way. He begins by accurately stating state hedonic adaptation limits the duration of hedonic loss, but later sneaks in the additional stipulation that the process of adaptation is itself relatively costless. Sunstein never argues for this additional claim other than to say it is "not implausible." If Sunstein is wrong, however, and adaptation imposes significant costs, the legal system would have reason to disregard Jones's injury in the first place.

A similar sleight of hand is at work the way that U&L motivate their paper. Like Sunstein, U&L looks only at the (long-term) after measurements of happiness: "Most people agree that kidney failure and quadriplegia are highly undesirable...yet measures of welfare based

³⁴ *Id.* at 15.

³⁵ For a graphical representation of U.S. unemployment trends, see Unemployment Rate (Seasonally Adjusted), U.S. Department of Labor, http://data.bls.gov/PDQ/servlet/SurveyOutputServlet?request_action=wh&graph_name=LN_cpsbref3 (last visited April 25, 2008).

on experience utility might give very little negative value to these patently undesirable conditions.”³⁶ From this snap-shot approach, U&L conclude there is nothing undesirable about *becoming* paraplegic.

Even worse, when U&L set up their final hypothetical, they cut out the process of adaptation all together:

[I]magine for a moment that you are about to receive a below the knee amputation. Imagine also, for the purpose of argument, that you are convinced that such an amputation will have only negligible effects on your moment-to-moment mood. How much money would you pay to avoid the amputation? If you are like us, you will pay a large amount.³⁷

In this hypothetical, U&L are trying to create a stylized hypothetical in which adaptation is assumed to be costless. Although U&L are not claiming that adaptation is actually costless, the persuasiveness of their hypothetical relies on the plausibility of this claim. If adaptation is extremely costly in real life, then one’s intuitions about the case of costless adaptation seem less relevant or reliable.

By ignoring the costs of adaptation, Sunstein and U&L create the illusion of a paradox; if adaptation is costly, then it is no mystery why people would want to avoid injury. For example, suppose in the standard hypothetical we replaced the permanent arm-amputation with a single instance of arm-stabbing. In the amputation case, people are surprised that amputees are able to recover so well psychologically. But, in the case of a stabbing, even lay people expect to recover. That is, even without hedonic adaptation research, people do not expect a single knife wound to inflict serious long-lasting hedonic damage. But, you could ask the same question: *How much money would you pay to avoid being stabbed?* I would certainly pay a lot. By Sunstein’s logic, this should be puzzling: My long-term happiness post-stabbing will match my pre-stabbing

³⁶ Lowenstein & Ubel, *supra* note 2, at 9.

³⁷ *Id.* at 11.

happiness. Thus, I will not have suffered any hedonic loss! Or, following U&L's approach, we could stipulate that adaptation is costless, *i.e.* that even right after the stabbing, the injury has no impact on my hedonic wellbeing. In that case, I would say that I have no intuition about the case. My aversion to stabbing is premised on the very belief that stabbing is painful!

Hedonic research has revealed something counter-intuitive: permanent physical injury does not necessarily impose permanent hedonic loss. So, losing a limb is not as bad as one might expect. That said, it is still pretty bad. Like serious non-permanent injuries, amputation is painful and recovery difficult. So, hedonic adaptation research should make people less averse to amputation. But, it should not make people indifferent to amputation as Sunstein and U&L seem to imply.

ii. Juries are likely not overcompensating catastrophic losses

The previous section argued that hedonic adaptation does not completely vitiate tort victims' claims for hedonic loss; the process of adaptation itself imposes hedonic losses worthy of compensation. Sunstein and U&L are right, however, that hedonic adaptation limits the duration of hedonic loss. Along this dimension, juries are likely to overestimate the extent of injury. At first approximation, it looks like "in the award of damages, the legal system is likely to be showing a systematic bias [toward over-compensation]"³⁸ In this section, however, I will argue that this worry might be overstated. Although juries may over-estimate the duration of a victim's suffering, they are also likely duration-insensitive in their willingness to compensate. These two psychological biases cut against each other, making it difficult to predict the ultimate direction of bias. Even if juries overcompensate, they will not be doing so to the degree feared by Sunstein and U&L. Taking duration-insensitivity into consideration, the real problem of tort compensation seems to be under-compensation of long-term harms.

³⁸ Sunstein, *supra* note 1, at 13.

Duration-insensitivity in one's willingness-to-pay (WTP) for adverse health conditions is part of the general phenomenon of scale-insensitivity in contingent evaluation. Contingent valuation surveys ask subjects to report their WTP for various goods for which market prices are difficult to ascertain. As Kahneman and Knetsch demonstrated in *Valuing Public Goods: The Purchase of Moral Satisfaction*, contingent evaluation does not always produce rational or robust values.³⁹ In particular, their surveys showed that people's WTP for public goods were not proportionately or sometimes even rationally related to scale. For example, people were willing to pay more to replant trees in British Columbia than they were to pay to replant trees in all of western Canada.⁴⁰ In a less dramatic example, people's willingness to pay to reduce damage in all of eastern Canada was only a few times more than their willingness to pay to reduce damage in Muskoka, Ont.⁴¹ Thus, one feature of contingent evaluation is that there is no linear (or even necessarily positive) relationship between WTP and scale of benefit.

In the health context, contingent evaluation surveys ask subjects to report their WTP to avoid certain bad health outcomes. Such surveys show insensitivity to a variety of factors including the duration of the health condition. In a typical study, the surveyor asks the subject how much they would pay to avoid suffering X days of Y condition. For example, one telephone survey asked the following question:

Assume that money can help you to avoid [symptom] and you can pay for it from your own pocket, that is, you will have less to spend on other goods or services. Please take a moment to think about the pain and suffering caused by [symptom] and the medical expense you will spend to relieve it." (a five second break) If you would experience one day of [symptom] tomorrow, what is the maximum amount you are willing to pay to avoid *one* day of it?⁴²

³⁹ Daniel Kahneman & Jack L. Knetsch, *Valuing Public Goods: The Purchase of Moral Satisfaction*, 22 J. ENVIRON. ECON. & MGMT 57 (1992).

⁴⁰ *Id.* at 65 tbl. 2.

⁴¹ *Id.* (reporting that for Muskoka and eastern Canada respectively, the median WTP was 20 and 50, and the mean WTP was 40.91 and 214.55)

⁴² Raymond Y.T. Yeung, Richard D. Smith & Sarah M. McGhee, *Willingness to Pay and Size of Health Benefits: An Integrated Model to Test for 'Sensitivity to Scale,'* 12 HEALTH ECON. 791 (2003).

The question is then varied for various symptoms (coughing, shortness of breath, and congested throat) and various time periods (three and seven days).⁴³ Looking at both external (between-sample) comparisons and internal (within-subject) comparison, researchers concluded that WTP is not proportional to duration.⁴⁴

Another way to measure the effect is to indirectly compute WTP based on respondents' graded-pair and discrete choices. In each choice, researchers present respondents with a series of alternative health scenarios. Each health scenario is described with along several dimensions, such as symptoms, duration, severity (in terms of limits on life activities) and cost. In the graded-pair choice condition, respondents evaluate two scenarios and rate their preference on a 7-point scale from "A is much better than B" through "B is much better than A." In the discrete choice condition, respondents evaluate three scenarios and simply indicate which is the their most preferred state. In a 2000 study, researchers incorporated respondent choices of both types and estimated WTP for each type of symptom, severity, and duration.⁴⁵ As with the direct state preference study, researchers in this study found imperfect sensitivity to duration. Even for pain that requires in hospital stays, the WTP for a 10-day stay was only roughly twice as much as the 1-day stay:

<i>Pain</i>	<i>In Hospital</i>
1 day	510 (426/611)
5 day	827 (665/1015)
10 day	1002 (791/1240)

⁴³ *Id.* at 794 tbl. 2.

⁴⁴ *Id.* at 795.

⁴⁵ F. Reed Johnson, Mellisa Ruby Banzhap & William H. Desvousges, *Willingness to Pay for Improved Respiratory and Cardiovascular Health: A Multiple-Format Stated-Preference Approach*, 9 HEALTH ECON. 295 (2000).

For all the symptoms, WTP does increase with duration, but marginal WTP drops off with increasing duration.

Finally, a meta-regression of studies recently confirmed this overall trend. In this 2006 paper, researchers compiled thirteen WTP studies.⁴⁶ To compare the studies against each other, researchers rated the health outcomes tested in each studies by the number of Quality Adjusted Life Years (QALY) lost. They then converted WTP results in terms of dollars paid per QALY gained. Using these thirteen studies, researchers ran a regression model with dependent variables of duration, severity, and various controls including characteristics of the study population, valuation method, study design, and publication outlet.⁴⁷ The meta-regression confirmed that “WTP increases less than proportionately with respect to changes in duration.”⁴⁸

Interestingly, George Lowenstein currently has a paper with Dan Ariely and Drazen Prelec in which they also report a duration-insensitivity effect.⁴⁹ In this paper, authors are primarily investigating the sensitivity of willingness-to-accept (WTA) to anchoring effects. Researchers run several experiments to elicit respondents’ WTA for listening to unpleasant noise lasting 10 seconds, 30 seconds, and 60 seconds. Unlike studies previously discussed, respondents get to fully experience each condition, sometimes multiple times, before making their choices. So the study has the advantage that respondents are not reporting WTA for hypothetical experiences but actual experiences that they themselves have just experienced. Even for such minor differences in scale (10 seconds versus 60 seconds), WTA does not increase proportionate to duration.⁵⁰ Authors, however, do not note the insensitivity to duration. Quite the opposite,

⁴⁶ George Van Houtven et al, *Valuing Avoided Morbidity Using Meta-Regression Analysis: What Can Health Status Measures and QALYs tell us about WTP?*, 15 HEALTH ECON. 775 (20006).

⁴⁷ *Id.* at 783.

⁴⁸ *Id.* at 787.

⁴⁹ Dan Ariely, George Lowenstein & Drazen Prelec, *Coherent Arbitrariness: Duration-Sensitive Pricing of Hedonic Stimuli Around An Arbitrary Anchor*, available at http://papers.ssrn.com/sol3/papers.cfm?abstract_id=243109.

⁵⁰ See, for example, results at page 17.

authors point to the fact that WTA increases with duration as a sign of duration-sensitivity. Hence, the title: *Coherent Arbitrariness: Duration-Sensitive Pricing of Hedonic Stimuli Around An Arbitrary Anchor*. The Ariely, Lowenstein & Prelee paper underscores the point that duration-insensitivity is a misleading term. A more accurate description of the phenomenon would be imperfect duration sensitivity, since WTP and WTA both do increase somewhat with duration.

What does imperfect duration sensitivity show us about jury compensation? If juries are not proportionately sensitive to the duration of victims' pain and suffering, then the fact that they are misled about the duration may not matter in the end. The punch-line to U&L hypothetical is that even knowing that amputation does not cause long-term suffering, we would still pay a lot to avoid it. This is a claim about WTP. But if our WTP for one year's worth of pain is not that much smaller than our WTP for ten years of pain, then we should not be surprised. Even if all we cared about were strictly hedonic losses, we should not expect people to discount their WTP proportionate to the duration of the suffering.

So, whether or not juries overcompensate depends on which bias is stronger – our tendency to overpredict duration or our tendency to underpay for duration. The WTP and WTA studies cannot give us precise estimates; these studies compare short term health conditions lasting, for example, from one day to ten days. As such, they are not great proxies for how juries compare pain and suffering that last from one year to ten years. At least we can say, however, with the counter-veiling bias, the worry about overcompensation is not as large as Sunstein and U&L may believe. Rather, the real worry should be about undercompensation for injuries that do in fact cause long-term harm, especially those for which we underestimate the duration to begin with.

In sum, hedonic wellbeing does not deny that victims of catastrophic injury should be compensated. Even with hedonic adaptation, tort victims will be compensated for the significant short-term harms that they suffer. Moreover, juries are not as likely to overcompensate for this loss as Sunstein and U&L fear. Therefore, Sunstein and U&L have made the problem appear bigger than it actually is. Hedonic adaptation does not strip the tort system of its theoretical underpinnings. Nonetheless, Sunstein and U&L still offer interesting proposals for alternative theories of tort compensation. In the next two Parts, I will examine each in turn and argue why neither alternative is appropriate.

Desire Satisfaction is Inappropriate When Underlying Preferences are Unreliable

U&L appeal to victims' own preferences for their pre-injury state to justify extra-hedonic compensation. Authors acknowledge that people are, in general, biased in their predictions and recollections of happiness. These biased estimates may be driving their preferences. If so, the victims' preferences present no challenge to hedonic wellbeing theories. U&L argue, however, that there is an alternative explanation, namely that the preferences reflect real preferences that are independent of hedonic experience.

When theory conflicts with intuition, the general dispute resolution method is reflective equilibrium. In this case, however, reflective equilibrium does not provide clear guidance. Everyone agrees that reliable moral intuitions should carry strong deliberative weight. But what about suspect moral intuitions? In response to doubts of reliability, U&L argue that at least *their* preferences are reliable because they are aware of potential biases. They have the same preferences even though they know that their long hedonic wellbeing will recover. In pointing out that their preferences are robust, U&L assume that rational deliberation can remove the preexisting bias. Empirical evidence, however, casts doubt on this assumption.

Preferences about tort compensation are morally loaded. In arguing that their own preferences are unbiased, U&L rely on a rational model of moral deliberation. If this model is accurate, U&L can freely de-bias themselves of rational errors in forming their moral preferences. Recent psychological research, however, suggests that moral intuitions in fact precede rational deliberation. At the very least, they occur in tandem. If these models are accurate, moral intuitions may work like visual illusions; knowing their logic does not make them go away. Consequently, the fact that U&L have the same preferences after rational reflection is not evidence of their reliability. More likely, it is evidence that moral preferences cannot be de-biased

That is not to say that policymakers should disregard individual preferences altogether. Rather, policymakers should weigh preferences according to their reliability. Preferences that are not morally loaded and are not subject to prima facie biases should be given the greatest weight. Within morally loaded preferences, those that are less personal and visceral should be given greater weight. Unfortunately, the case of pain and suffering damages for gruesome injuries is the worst case to rely on victim preferences.

i. An Intuition and An Error Theory

U&L believe that people prefer healthy states over unhealthy states even independent of any hedonic differences. In support, they point to the fact that “people experiencing these [unhealthy] conditions, despite reporting levels of mood and well-being that are similar to healthy persons, also report a willingness to pay large sums of money, or to make costly tradeoffs, to restore their lost function.”⁵¹ Unlike the rest of the population, people actually living in with amputations or colostomies have direct knowledge about these hedonic experiences. They are not subject to the mis-prediction biases discussed in the last part; they know they are

⁵¹ Lowenstein & Ubel, *supra* note 2, at 9.

still fairly well off hedonically. Their preferences reveal, therefore, that hedonic experience does not capture everything there is to value.

As U&L acknowledge, however, there are two other biases that potentially explain away these preferences. First, people mis-predict how happy they would be if they were to reverse their health condition: “Although people with chronic conditions report similar levels of happiness to those in good health, when asked to estimate what their happiness would be if they did not have their health condition, patients tend to predict that they would be happier than they are.”⁵² Second, people mis-remember how happy they were before their health condition. For example, “[t]hose with colostomies, thinking back to a time when they did not have a colostomy, ‘remembered’ being significantly happier than they were currently.”⁵³ If people either mis-predict future unhappiness or mis-remember past happiness, they would have purely hedonic (albeit erroneous) reasons for wanting to reverse their health condition. As U&L put it, “[s]uch preferences could, of course, [just] be misinformed.”⁵⁴

This is not, however, U&L’s preferred interpretation: “[A]nother plausible interpretation is that both groups [healthy and unhealthy] are attempting to express the same thing: chronic illness is undesirable.”⁵⁵ If the preferences are not tracking hedonic wellbeing, then biases about hedonic experience are irrelevant; victims simply have non-hedonic preferences.

Unfortunately, U&L do not provide any criteria on which to evaluate these competing interpretations. In fact, there are no agreed upon criteria of resolving such conflicts within moral theory. In the next section I will discuss the general method of reflective equilibrium and its

⁵² *Id.* at 9.

⁵³ *Id.* at 10.

⁵⁴ *Id.* at 11.

⁵⁵ *Id.* at 9.

inability to produce a solution in this case. I will argue that reflective equilibrium cannot be applied without a deeper understanding of the process of intuition formation.

ii. Reflective Equilibrium Provides No Determinate Answer

Our intuitions, as interpreted by U&L, conflict with hedonic theories of happiness. The method of reflective equilibrium aims to produce coherence between intuitions and theory.⁵⁶ In the process of reflective equilibrium, we make trade-offs between modifying our theory (at the risk of being ad hoc) and rejecting our intuition (at the risk of offending common sense). The end result is a combination of “judgments, principles, and theories [that] we are un-inclined to revise any further because together they have the highest degree of acceptability or credibility for us.”⁵⁷

Although theorists agree on the goal of reaching reflective equilibrium, there is no consensus the process of achieving it. Kwame Appiah tackles this tough issue in his new book, *Experiments in Ethics*.⁵⁸ Given any difficult trade-off between theory and common sense, there will be proponents on both sides: “A theory may be applauded for being revisionary, for showing us the error of our quondam judgments – or rejected precisely because it defies common sense.”⁵⁹ If the two competing theories are equally coherent, reflective equilibrium is essentially indifferent between them. Thus, as a process of theory building, reflective equilibrium is vastly indeterminate. As Appiah puts it: “Indeed, one could be forgiven for thinking that reflective equilibrium is really another name for the problem, rather than a solution to it.”⁶⁰

The general conflict between theory and intuition is complicated further when the intuition is itself suspect. The simpler case is when you have a moral intuition that is clear,

⁵⁶ For a history of reflective equilibrium in philosophy, see Norman Daniels, *Reflective Equilibrium*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY, <http://plato.stanford.edu/entries/reflective-equilibrium/> (2003).

⁵⁷ *Id.*

⁵⁸ KWAME ANTHONY APPIAH, *EXPERIMENTS IN ETHICS* (2008).

⁵⁹ *Id.* at 76.

⁶⁰ *Id.* at 78.

widely-shared, and non-suspect (in terms of reliability). Theorists agree that these intuitions carry great deliberative weight; failure to account them is a serious mark against the strength of any moral theory. Appiah quotes Thomas Reid’s formulation of the basic rule: “A consent of ages and different nations of the learned and vulgar, ought, at least to have great authority, *unless we can show some prejudice as universal as that consent is, which might be the cause of it.*”⁶¹ Later, Appiah himself provides a similar formulation: If something seems intuitively wrong and *you have no special knowledge that suggests your moral intuition is distorted* you shouldn’t do it.⁶² But what about when you do have special knowledge that your intuition could be distorted? Cognitive biases seem to be exactly the type of “universal prejudice” that defeats the presumption in favor of shared moral intuitions.

Ultimately, Appiah is unclear about what deliberative weight, if any, suspect intuitions should carry. Appiah concludes that an understanding of cognitive biases is relevant to our evaluation of intuitions: “By illuminating odd features of our evaluations, showing where judgments about cases of a certain form diverge from our other, considered judgments, that research can be a useful adjunct to our intuitions...”⁶³ Appiah does not, however, clarify in what sense the research is “useful.” Do cognitive biases merely negate the presumption in favor of intuitions? In that case, our epistemological stance should be one of suspension of belief.⁶⁴ Alternatively, do cognitive biases actually reverse the presumption in favor of intuition? And if so, is that presumption defeasible?

Perhaps Appiah could not be more specific because the answers differs across types of intuitions. In order to understand how cognitive biases impact the reliability of moral intuitions,

⁶¹ *Id.* at 74.

⁶² *Id.* at 113.

⁶³ *Id.* at 115.

⁶⁴ For an argument in favor of mere suspension of belief, see Walter Sinnott-Armstrong, *Framing Moral Intuition*, in 2 MORAL PSYCHOLOGY 47, 73 (Walter Sinnott-Armstrong ed. 2008).

we need to know what role cognitive biases play in their formation. Specifically, we need to know whether rational reflection can correct for the effects of cognitive biases. In the next section, I present four theories of moral intuition.⁶⁵ Under the Kantian theory, rational reflection can correct for cognitive biases. If this is the right theory, then cognitive biases pose no threat to the reliability of moral intuitions. Any intuition that survives informed rational reflection is still reliable. Under the other three Humean theories, rational reflection may not be able to correct for cognitive biases. Under some Humean theories, therefore, cognitive biases create indefeasible presumptions against intuition. In the final section, I will argue that the types of intuitions U&L rely on likely fall under one of the Humean models. Therefore, we should not rely on these intuitions in the formation of moral theory or public policy.

iii. Humean Models of Moral Judgment Cast Doubt on the Reliability of Certain Intuitions

A preliminary note on some confusing terminology: So far, I have been using moral “intuition” to mean judgments of folk morality. Such judgments are “intuitions” in the sense that they are pre-theoretic. In this section, I will be exploring different cognitive theories about how such judgments form. In this context, theorists use moral “intuition” to refer to the pre-conscious moral response, *i.e.* the sudden consciousness that appears without any thought at all. In contrast, the moral response that solidifies after conscious reflection is called a moral “judgment.” So for the purposes of this section, I will follow the convention of calling pre-conscious moral responses “intuitions” and conscious, reflective moral responses “judgments.”

There are two approaches to understanding moral judgment, epitomized by the great philosophers Immanuel Kant and David Hume. Crudely speaking, Kant argued that moral

⁶⁵ I am roughly adopting the topology of Marc D. Hauser, Liane Young & Fiery Cushman, *Reviving Rawls’s Linguistic Analogy: Operative Principles and the Causal Structure of Moral Actions*, in 2 MORAL PSYCHOLOGY, *supra* note 64, at 107. The major difference is that I consider all three non-Kantian models Humean in the broad sense that all three are anti-rationalist, whereas Hauser, Young, and Cushman identify only the Social Intuitionist model as strictly Humean in the sense that emotion precedes and drives judgment.

judgment is rational, driven by our unconstrained capacity for moral reasoning. Hume, in contrast, argued that both moral judgment and moral reasoning are driven by non-rational psychological forces such as affect and emotion.

To be clear, Kant did not deny that our actions are often driven by emotion rather than reason; he recognized that humans do not always act in accordance with their moral judgment. Kant did emphatically deny, however, that moral judgments themselves are driven by emotion. For Kant, moral judgments are grounded “exclusively on principles that are inherent in and revealed through the operations of reason.”⁶⁶ Therefore, the operations of reason must be unconstrained by any forces external to it, including natural, non-rational, psychological forces: “The concept of a rational will...[is] a will that does not operate through the influence of factors outside of this responsiveness to reasons. For a will to be free is thus for it to be physically and psychologically unforced in its operation.”⁶⁷ It is in this sense of absolute freedom from constraint that Kant believed that “rational human wills are autonomous.”⁶⁸

From a philosophical perspective, Kant’s moral theory mandated his rationalist model of moral judgment. For Kant, moral laws do not derive their binding force from an external source, such as God. Rather, moral laws bind us because only we impose them on ourselves. Like a democracy, we are self-governing in the sense that we bind ourselves through law. Moral laws have legitimacy only to the extent that we have the capacity to self-govern, that is, to the extent that we have free rational wills. Consequently, “[t]he fundamental principle of morality — the CI [Categorical Imperative] — is none other than the law of an autonomous will.”⁶⁹ Therefore, if

⁶⁶ Lara Denis, *Kant and Hume on Morality*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY (2008), <http://plato.stanford.edu/entries/kant-hume-morality/>.

⁶⁷ *Id.*

⁶⁸ *Id.*

⁶⁹ Robert Johnson, *Kant’s Moral Philosophy*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY, <http://plato.stanford.edu/entries/kant-moral/> (2008).

moral judgment were constrained by non-rational forces, moral laws would lose their legitimacy and binding force.

But from an empirical perspective, how did Kant know that our rational wills are in fact autonomous? Kant himself relied on the transcendental method: “[Kant’s] empirical method for doing psychology...was introspection.”⁷⁰ Kant’s followers in psychology, however, have sought to confirm the rationalist theory experimentally. Most notably, Lawrence Kohlberg looked for evidence within cognitive development.

Kohlberg followed Jean Piaget in studying how our faculty for moral reasoning develops from infancy into adulthood. Kohlberg developed a taxonomy of moral reasoning and argued that people universally shift from early stages (*e.g.*, instrumental reasoning) to later stages (*e.g.*, abstract reasoning regarding universal ethical principles) as they develop.⁷¹ For example, “younger children base their moral judgments more on consequences, whereas older children base their judgments on intentions.”⁷² Kohlberg’s experimental method involved probing subjects’ moral evaluations of ethical dilemmas, including the famous case of “Heinz Steals the Drug.”⁷³ The focus of Kohlberg’s interviews was not the content of the moral judgment but rather the reasoning behind them.

⁷⁰ Andrew Brook, *Kant’s View of Mind and Consciousness*, in STANFORD ENCYCLOPEDIA OF PHILOSOPHY, <http://plato.stanford.edu/entries/kant-mind/> (2004).

⁷¹ For a succinct review of Kohlberg’s six-stage model, see Lawrence Kohlberg, *The Claim of Moral Adequacy of a Highest Stage of Moral Judgment*, 70 J. PHILO. 630, 631 tbl.1 (1973).

⁷² W.C. CRAIN, THEORIES OF DEVELOPMENT 118-36 (1985).

⁷³ The ethical dilemma is described as follows: “In Europe, a woman was near death from a special kind of cancer. There was one drug that the doctors thought might save her. It was a form of radium that a druggist in the same town had recently discovered. The drug was expensive to make, but the druggist was charging ten times what the drug cost him to make. He paid \$200 for the radium and charged \$2,000 for a small dose of the drug. The sick woman’s husband, Heinz, went to everyone he knew to borrow the money, but he could only get together about \$ 1,000 which is half of what it cost. He told the druggist that his wife was dying and asked him to sell it cheaper or let him pay later. But the druggist said: ‘No, I discovered the drug and I’m going to make money from it.’ So Heinz got desperate and broke into the man’s store to steal the drug-for his wife. Should the husband have done that?” Lawrence Kohlberg, *The Development of Children’s Orientation Toward a Moral Order: Sequence in the Development of Moral Thought*, 6 VITA HUMANA 11, 19 (1963).

Kohlberg's interview studies established that moral reasoning *accompanies* moral judgment. While this finding is consistent with either Kantian or Humean models, Kohlberg adopted the Kantian interpretation, *i.e.* that moral reasoning *drives* moral judgment. Furthermore, Kohlberg saw the natural step-wise development of moral reasoning as a path toward increasing rationality.⁷⁴ That is, Kohlberg argued that the later stages of more abstract deliberation constituted normatively better forms of reasoning than the earlier stages of more concrete, instrumental reasoning.⁷⁵

If Kant and Kohlberg have the right model, then moral judgments are rational although not infallible. Moral judgments are rational in that they are caused by reason not emotion. In Kohlberg's words "the moral is cognitive."⁷⁶ Moral judgments are not infallible, however, because human beings can err in their reasoning. Cognitive biases may introduce common errors in moral reasoning. But to the extent that human beings have the capacity to recognize such errors, moral judgments are not hopelessly unreliable. Once these fallacies are discovered, people can adjust their moral reasoning, and therefore correct their moral judgments.

A helpful analogy here is to logical proofs: People often get the wrong answer by making fallacious deductions. People do not, however, persist in believing the conclusion after they are conscious of the errors that produced it. If someone does persist in holding onto their prior conclusion, it can only be because (1) he disagrees on what constitutes proper reasoning, or (2) his rational capacity is limited such that he cannot recognize his own error. Outside these two possibilities, cognitive biases should not constrain moral judgment.

⁷⁴ See DANIEL L. LAPSLEY, MORAL PSYCHOLOGY 49-50 (1996) (analyzing Kohlberg's 1971 article, *From is to Ought: How to Commit the Naturalistic Fallacy and Get Away With it in the Study of Moral Developments*).

⁷⁵ *Id.*

⁷⁶ Lawrence Kohlberg, *From is to Ought: How to Commit the Naturalistic Fallacy and Get Away With it in the Study of Moral Developments*, in COGNITIVE DEVELOPMENT AND EPISTEMOLOGY 151, 230 (T. Mischel Ed. 1971).

The rationalist model, therefore, paints a fairly hopeful picture for the reliability of moral judgments. In contrast, Humean models are much more skeptical of the ability of reason to rehabilitate biased moral judgments. Hume famously claimed, “Reason is, and ought only to be the slave of the passions, and can never pretend to any other office than to serve and obey them.”⁷⁷ Hume recognized that humans engage in moral reasoning all the time. He argued, however, that the role of moral reasoning “is an auxiliary one.”⁷⁸ A strict Humean would argue that emotion, not reason, determines moral judgment. A weaker anti-rationalist claim, however, merely argues that reason does not have an unmediated connection to moral judgment, *i.e.* that non-rational forces play some causal role in this relationship. In this paper, I will refer to this weaker claim as “Humean” in the broad sense.

Recently, a wave of empirical work has come out in support of the anti-rationalists. Researchers use two complementary strategies to examine the role of emotion. On one side, studies show that manipulating ordinary people’s emotional responses has an impact on their moral judgments. For example, in a revealing experiment, Talia Wheatley & Jonathan Haidt selected arbitrary words and hypnotized subjects to feel a “flash of disgust” upon reading them.⁷⁹ Subjects were then presented with descriptions of moral transgressions. When asked to rate the hypotheticals, subjects made harsher moral judgments when the descriptions contained the selected words. The paper’s title says it all: *Hypnotic Disgust Makes Moral Judgments More Severe*. The study supports the weaker Humean position; while emotional responses were not the sole determinant, they did causally affect the content of people’s judgments.

⁷⁷ David Hume, *A Treatise of Human Nature* 297 (1739-40) (NuVision Publications ed. 2007).

⁷⁸ Denis, *supra* note 66.

⁷⁹ Talia Wheatley & Jonathan Haidt, *Hypnotic Disgust Makes Moral Judgments More Severe*, 16 *PSYCHOL. SCI.* 780 (2005).

On the other side, studies show that people impaired emotional functioning have atypical moral judgments. Michael Koenigs et al, for example, studied the moral judgments of people with ventromedial prefrontal cortex (VMPC) damage.⁸⁰ The VMPC is the brain region active during the normal generation of emotions: “Patients with VMPC lesions exhibit generally diminished emotional responsivity and markedly reduced social emotions.”⁸¹ Authors found that moral judgments within this population were substantively different from that of the general population. Specifically, those with VMPC lesions are more likely to endorse utilitarian solutions to ethical dilemmas. Since VMPC lesion patients preserve normal intelligence, logical reasoning, and knowledge of social and moral norms, differences in moral judgment can only be attributed to differences in their emotional functioning.⁸² VMPC lesion patients did not, however, differ from the general population in all their moral judgments: “patients’ judgements [sic] differed from comparison subjects’ only for the high-conflict personal moral dilemmas.”⁸³ Therefore, this study also supports the weaker Humean position; emotional responses play a causal role in some but not all moral judgments.

Taken together, both lines of research demonstrate that emotions affect how we think about moral dilemmas. As one commentator exclaims, “[T]hat emotions play a part is undebatable.”⁸⁴ Now, the only question is how they do so. Within this broadly Humean perspective, three distinct theories of moral judgment have emerged.

The theory closest to expressing Hume’s own views is Haidt’s Social Intuitionism.⁸⁵ Under this model, unconscious, emotional, processes produce moral judgment; moral reasoning

⁸⁰ Michael Koenigs et al, *Damage to the Prefrontal Cortex Increases Utilitarian Moral Judgments*, 446 NATURE 908 (2007).

⁸¹ *Id.* at 908.

⁸²

⁸³ *Id.* at 910.

⁸⁴ Hauser, Young & Cushman, *supra* note 65, at 121.

⁸⁵ See, e.g., Jonathan Haidt, *The Emotional Dog and Its Rational Tail: A Social Intuitionist Approach to Moral Judgment*, 108 PSYCHOL. REV. 814 (2001).

merely follows and justifies moral judgment: “The central claim of the social intuitionist model is that moral judgment is caused by quick moral intuitions and is followed (when needed) by slow, ex post facto moral reasoning.”⁸⁶ Building on Haidt, theorists such as Gerd Gigerenzer hypothesize that our unconscious moral responses are driven by evolutionarily ingrained moral heuristics.⁸⁷ According to Gigerenzer, heuristics are fast and frugal rules of thumb. Heuristics are evolutionarily advantageous because they exploit internal advantages (our unique abilities) and situational advantages (such as when the environment produces informational redundancy). But “[h]euristics are always second-best solutions, which describe what people do but do not qualify as guidelines for moral action.”⁸⁸ Similarly, as Sunstein puts it, “in many contexts, ...heuristics...lead to moral error on any plausible view of morality.”⁸⁹

Under social intuitionism, cognitive biases could be built into certain moral heuristics. If so, such biases are evolutionarily ingrained. Rational reflection has little opportunity to override their influence because reasoning comes only after moral judgments are already formed. In Haidt’s words, “People may at times reason their way to a judgment by sheer force of logic, overriding their initial intuition...However, such reasoning is hypothesized to be rare, occurring primarily in cases in which the initial intuition is weak and processing capacity is high.”⁹⁰ That is not to say that heuristics bind our action. Rational reflection can persuade us not to act on potentially biased judgments. Rational reflection usually cannot, however, alter our underlying moral instincts.

⁸⁶ *Id.* at 817.

⁸⁷ Gerd Gigerenzer, *Moral Intuition – Fast and Frugal Heuristics?*, in 2 MORAL PSYCHOLOGY *supra* note 64, at 1.

⁸⁸ *Id.* at 19.

⁸⁹ Cass R. Sunstein, *Fast, Frugal, and (Sometimes) Wrong*, in 2 MORAL PSYCHOLOGY, *supra* note 64.

⁹⁰ Haidt, *supra* note 85, at 819.

In contrast to Social Intuitionism, Dual Process Theory hypothesizes that both unconscious intuitive reasoning and conscious reflective reasoning direct moral judgment.⁹¹ Kahneman and Sunstein label these two modes of cognition System 1 and System 2 respectively. System 1 operates quickly and “proposes intuitive answers to judgment problems as they arise.”⁹² System 1 answers, however, are only provisional. Before moral judgments form, System 2 has the opportunity to “endorse, correct, or override” these proposals.⁹³ According to Dual Process Theory, neither system has complete control over moral judgment: “automatic and controlled cognitive operations compete for the control over overt responses.”⁹⁴

Dual Process Theory leaves unclear how conflicts between System 1 and 2 get resolved psychologically.⁹⁵ Joshua D. Greene et al hypothesize that System 1 has stronger influence in resolving moral conflicts that are personal in nature, whereas System 2 has stronger influence in resolving impersonal conflicts.⁹⁶ Authors emphasize, however, that the distinction between System 1 and 2 is a fuzzy one: “[T]he emotion/cognition distinction is real, but it is a matter of degree, and at present time, not very well understood.”⁹⁷ Thus, Dual Process Theory leaves open the possibility that cognitive errors introduced by System 1 can be corrected by System 2. The theory, however, cautions against complete reliance on System 2, especially under circumstances in which System 1 tends to have greater influence.

⁹¹ For an overview of the theory, see Daniel Kahneman & Cass R. Sunstein, *Cognitive Psychology of Moral Intuitions*, in *NEUROBIOLOGY OF HUMAN VALUES* 91 (Jean-Pierre P. Changuex et al eds. 2005) (positing that for moral intuitions, System 1 is based on the unique emotion of indignation).

⁹² *Id.* at 93.

⁹³ *Id.*

⁹⁴ *Id.*

⁹⁵ Hauser, Young & Cushman, *supra* note 65, at 115.

⁹⁶ Joshua D. Greene et al, *The Neural Bases of Cognitive Conflict and Control in Moral Judgment*, 44 *NEURON* 389 (2004) (relying on fMRI imaging to detect the relative dominance of System 1 and System 2 operations).

⁹⁷ *Id.* at 398.

Finally, the theoretical opposite of Dual Process Theory is the Moral Linguistic Theory, which posits that neither System 1 nor System 2 influences moral judgment.⁹⁸ As Marc D. Hauser, Liane Young, and Fiery Cushman argue, the process of moral judgment formation is entirely unconscious and neither affective or rational: “Prior to either emotion or deliberate reasoning, there must be some kind of unconscious appraisal mechanism that provides analysis of the causes and consequences of action.”⁹⁹ Essentially, moral judgment happens in a black box; the resulting moral judgment then triggers both our emotional and cognitive responses.

Inspired by Rawls,¹⁰⁰ Moral Linguist Theory analogizes moral capacity to linguistic capacity. According to Chomsky’s theory of linguistic acquisition, humans are endowed with innate language propensities that develop in childhood through appropriate exposure.¹⁰¹ As children, we learn to employ the principles of language unconsciously and effortlessly. Only by studying how people use language can we reverse engineer those linguistic principles. Moral Linguistics Theory hypothesizes that in an analogous way, humans are endowed with universal moral propensities, which we develop through socialization. As with linguistic principles, we can discover moral principles by studying people’s moral judgments. But those principles will not be consciously accessible to us even as we employ them.

According to Moral Linguistic Theory, moral judgments are influenced by innate propensities and childhood influences. Emotional or deliberative responses to any particular moral dilemma cannot influence our moral judgment about the case. These responses are effects not causes: “[e]motion and deliberative reasoning are not causally related to our initial moral

⁹⁸ See Hauser, Young & Cushman, *supra* note 65.

⁹⁹ *Id.* at 121.

¹⁰⁰ *Id.* at 114.

¹⁰¹ See, e.g., Noam Chomsky, *Cartesian Linguistics* (1966).

judgments.”¹⁰² Therefore, if there are biases or irrationalities functioning within the moral module, they are inaccessible and uncorrectable.

In sum, there are four theories of moral judgment. Each assigns a different role to moral reasoning:

	<i>How do moral judgments form?</i>
Rationalist	Reasoning → Judgment → Emotion
Social Intuitionism	Emotion → Judgment → Reason
Dual Process Theory	Emotion + Reason → Judgment
Moral Linguistic Theory	Judgment → Emotion + Reason

Consequently, each provides a different assessment of our ability to correct for cognitive biases. Under rationalist theories, reason is always free to correct errors as long as those irrationalities are known. Under Dual Process Theory, reason can sometimes correct for errors, most likely when faced with more impersonal moral choices. Finally, under either Social Intuitionism or Moral Linguist Theory, reason cannot correct for errors because reason follows judgment.

Returning to the question of reflective equilibrium, whether we can rely on potentially biased moral intuitions depends on which theory of moral judgment is correct. If we believe the Rationalist model, then the fact that moral judgments survive reflective reasoning shows that the judgment is picking up on a real moral difference rather than some spurious cognitive error. Moral theory, must meet the demands of intuition and figure out what that moral difference is. On the other hand, if either Social Intuitionism or Moral Linguistic Theory are correct, then the discovery of irrational biases creates irrefutable presumption against the moral judgment. More

¹⁰² *Id.*

reflective reasoning cannot improve the reliability of our moral judgments. Finally, if Dual Process Theory is correct, then the reliability of any given moral judgment will vary by degree. Cognitive biases introduce a presumption against intuition, but that presumption may be rebutted in some circumstances by better quality reasoning.

iv. U&L's Intuitions are Likely Unreliable

The obvious question, then, is which model is right? Or if there is no universal model, which model explains the intuition at hand? U&L claim that their moral intuitions are reliable because they persist after rational reflection with conscious awareness of the potential cognitive errors. Therefore, U&L are implicitly assuming that we are operating under either a Rationalist or Dual Process Theory model. As I argued in the previous section, empirical research reveals that non-rational forces play at least some causal role in moral judgment. So it is unlikely that the pure Rationalist model is correct. Therefore, U&L are justified in relying on their intuitions only if Dual Process Theory is the right Humean theory and the specific intuition is one in which System 2 dominates System 1. In this section, I will argue that this condition is unlikely to hold.

U&L do not offer explicit evidence in support of the reliability of their intuition. From the form of their argument, U&L seem to be relying on introspection. U&L acknowledge why one might be skeptical about victims' intuitions; their intuitions may be based on inaccurate hedonic forecasting. Ultimately, they do not hinge their argument on other people's intuitions, but instead on their own intuitions: "*If you are like us, you will pay a large amount [to avoid amputation].*"¹⁰³ Perhaps U&L find their own intuitions more persuasive because they believe they have privileged access to them. From introspection, they can confirm that at least their own intuitions are not motivated by concerns of hedonic wellbeing.

¹⁰³ *Id.* at 11.

Unfortunately, U&L’s introspective evidence does not actually shed light on the question. Introspection cannot reveal whether moral reasoning precedes or follows moral judgment. Either way, we will experience moral reasoning as justifying our moral judgments. As Walter Sinnott-Armstrong argues, “The fact that our moral intuitions seem justified does not show that they really are justified.”¹⁰⁴ The experience of moral justification is what Haidt calls the wag-the-dog illusion: “We believe that our own moral judgment (the dog) is driven by our own moral reasoning (the tail).”¹⁰⁵ U&L could very well be right in this case, but their internal experience does not constitute a valid check. Sinnott-Armstrong analogizes the epistemic situation to a case of good and bad thermometers: “Maybe we get lucky and our personal moral thermometer is one of the ones that works fine, but we are still not justified in trusting it, if we know that lots of moral thermometers do not work, and we have no way of confirming which ones do work.”¹⁰⁶ Without more evidence, we cannot rule out Social Intuitionism or Moral Linguistic Theory models. Therefore, U&L are not justified in assuming that the Dual Process Theory applies, *i.e.* that reflective reasoning has an opportunity to impact on judgment at all.

Furthermore, even if we grant that the Dual Process Theory is the right model, U&L are still not justified in relying on their intuitions. As discussed above, under Dual Process Theory, System 1 and 2 compete for dominance. Given potential biases introduced by System 1, U&L need System 2 to override System 1. This is more likely when the moral conflict is impersonal (*e.g.*, switching the trolley tracks) and less likely when the conflict is “up close and personal” (*e.g.*, pushing the fat man off the bridge).¹⁰⁷ As Hauser, Young, and Cushman explain, System 1 is likely to dominate when “a highly aversive action is in conflict with the generation of a

¹⁰⁴ Walter Sinnott-Armstrong, *Framing Moral Intuition*, in 2 MORAL PSYCHOLOGY, *supra* note 64, at 48.

¹⁰⁵ Haidt, *supra* note 85, at 823.

¹⁰⁶ Sinnott-Armstrong, *supra* note 104, at 71.

¹⁰⁷ Joshua D. Greene et al, *An fMRI Investigation of Emotional Engagement in Moral Judgment*, 293 SCIENCE 2105 (2001).

significant utilitarian outcome, and the action involves personal contact with another.”¹⁰⁸ The stylized hypothetical that U&L set up fits this mold. Withholding compensation from the victim of a serious tort is “aversive” (although concededly not as aversive as pushing someone off a bridge). Paying the victim, on the other hand, conflicts with utilitarian principles since U&L stipulate that the tort causes absolutely no hedonic harm. Finally, the moral dilemma is personal; the question is not whether some fat man should lose a leg, but whether you yourself should lose a leg! It is difficult to get more up close and personal than self-amputation.

U&L have created their hypothetical to illicit a strong moral reaction. In doing so, however, they have also made it likely that System 1 will dominate System 2. Given that we have evidence to believe that System 1 is biased, U&L are not justified in relying on their System 1-driven moral intuition. More broadly, theorists and policy makers should weigh moral intuitions in accordance to their likely reliability. This does not mean disregarding our moral intuitions altogether. But it does mean avoiding potentially biased intuitions, especially those that are emotionally loaded.

This safeguard, however, still leaves plenty of room to challenge hedonic wellbeing theories. For example, another well known anti-hedonic intuition is the preference for leaving choices open.¹⁰⁹ As John Tierny explains in a recent *New York Times* article, “In a series of experiments, hundreds of students could not bear to let their options vanish, even though it was obviously a dumb strategy...”¹¹⁰ In the article experimenter, Dan Ariely, admits that even he himself cannot bear to close doors in his own life.¹¹¹ One interpretation is that our preferences are irrational; keeping choices open actually makes people hedonically worse off. Under this

¹⁰⁸ Hauser, Young & Cushman, *supra* note 65, at 138.

¹⁰⁹ See, e.g., John Tierny, *The Advantages of Closing a Few Doors*, N.Y. TIMES, Feb. 26, 2006 (discussing DAN ARIELY, PREDICTABLY IRRATIONAL: THE HIDDEN FORCES THAT SHAPE OUR DECISIONS (2008))

¹¹⁰ *Id.*

¹¹¹ *Id.*

interpretation governmental policies should limit choices, even though this runs against most people's preferences. Another interpretation, however, is that people have genuinely non-hedonic preferences. Under this interpretation, the government ought to satisfy people's desires for maintaining maximal choice. The argument structure in this case is exactly the same as U&L's. But the argument for maintaining choice is more credible because the underlying intuition is abstract and impersonal (literally involving symbolic doors).¹¹² Therefore, there may well be other spheres in which desire fulfillment is the best theory of happiness. My argument here is only that U&L have not made a strong case for it in the context of tort liability.

To be clear, I am not arguing that U&L are necessarily wrong about their moral intuition. Rather, I argue only that along the range of justified reliance to unjustified reliance, U&L's intuition likely falls toward the unjustified end. Given that the existing hedonic theory already provides substantial grounds for compensating tort victims (as argued in the previous part), it would be unwise to adopt U&L's radical policy changes upon the strength of their suspect intuition.

Objective List Theories are Inappropriate for Retroactive Tort Compensation

Instead of appealing to victims' preferences, Sunstein appeals to objective list theories of capability to justify compensation. Some capabilities, Sunstein argues, are valuable whether or not you want them. To be successful, Sunstein and U&L must overcome the basic challenge of objective list theories: paternalism.

The objective list theorist must produce a principled basis on which to distinguish between potential human ends. Aristotle did so by relying on his overarching metaphysical theory. According to Aristotle, all things from the elements to the stars aim toward achieving the eternal and divine. Human functions that bring us closer to the eternal produce more objective

¹¹² To clarify, however, I am not claiming that the argument is ultimately successful.

wellbeing; to be fully human is to exercise these functions to the fullest extent. Unfortunately, Aristotle's metaphysical assumptions are unappealing to a modern audience.

Modern objective list theorists, therefore, must come up with new strategies to combat paternalism. The Capabilities Approach does so by limiting the scope of the theory. The Capabilities Approach does not aim to specify what it means to be fully human. Rather, the Capabilities Approach aims to specify what it means to be minimally human. That is, what are the functions without which a thing can not be called "human"? The Capabilities Approach lists ten functions and justifies the list on the basis of consensus. Even if we cannot agree on the ultimate ends of human life, we can at least agree on its minimal requirements. Finally, the Capabilities Approach emphasizes that the theory does not prescribe any individual course of action; it only demands that governments provide the resources necessary for individuals to act on these capabilities.

Thus, Capabilities Approach avoids paternalism by being limited in scope, consensus based, and prospective. The structure of tort compensation, however, negates each of these basic conditions. Most torts are not serious enough to make a victim sub-human. To compensate for the value of one leg, for example, the tort system requires a comprehensive and detailed list of what it means to be fully human. Consequently, the list cannot be truly consensus based, but can at best be majority based. Finally, the list cannot be used in prospective ways; the government cannot bring back leg capabilities. The list can only be used in a retrospective way, compensating a victim for his irreversible intrinsic loss.

Without the theoretical underpinning of the Capabilities Approach, applying an objective list theory is dangerously paternalistic. Instead of paying victims for the cost of adaptation, we would be paying them for the permanent loss of their humanity. This policy poses the serious

risk of branding victims as less than fully human. Before jumping on board with Sunstein's recommendations, therefore, we must consider whether the value of compensating non-hedonic is worth the potential risk.

i. Sunstein's Appeal to Objective List Theories

Sunstein explicitly appeals to objective list theories, and in particular to the Capabilities Approach. Recall that Sunstein's primary hypothetical involved the unlucky Jones who loses a leg but suffers no hedonic damage. Sunstein states: "What Jones has lost is a capability."¹¹³ In a footnote, Sunstein clarifies: "I am not using the idea of 'capabilities' in the same sense as Sen and Nussbaum, but my use belongs in the same general family, focusing as it does on the capacity to function, rather than subjective mental states."¹¹⁴ Other than referencing Sen and Nussbaum, Sunstein does not provide a definition of capability. What Sunstein seems to have in mind is list of physical and mental abilities as measured against the baseline of average or "normal" functioning. For example, Jones has lost a capability because "[h]e cannot walk on his leg; he certainly is unable to run. He is unable to engage in many activities that he *used to take for granted*."¹¹⁵

Similarly, Sunstein gives no account of why capabilities are valuable. He alludes to two potential rationales, neither of which are compelling. First, he argues that people prefer having capabilities: "[M]ost people would be willing to pay significant amounts to avoid a loss of a capability, even if they could be persuaded that the loss would inflict no hedonic harm."¹¹⁶ This rationale simply converts his objective list argument into a desire fulfillment argument. Sunstein

¹¹³ Sunstein, *supra* note 1, at 15.

¹¹⁴ *Id.* at n.14.

¹¹⁵ *Id.* (emphasis added).

¹¹⁶ *Id.* at 15.

himself is skeptical of the U&L argument discussed in the previous section.¹¹⁷ Furthermore, the rationale obscures a crucial (and potentially unattractive) feature of his own argument: According to objective list theories, capabilities are valuable even if you would prefer not to have them. In fact, if you think your own life is richer and more fulfilling without the capability, you are just objectively wrong.

Second, Sunstein argues that capabilities are valuable because they allow for diversity of subjective experience. Hedonic wellbeing theories do not capture the value of having experiences that are “numerous and qualitatively distinct.”¹¹⁸ This diversity of experience is only possible if you possess the requisite capabilities. This argument is also misleading. Is Sunstein really arguing that emotional variety is *per se* valuable? If so, then schizophrenics live the most valuable lives of all. Capabilities are not valuable because they expand the range of possible experiences. Rather, capabilities are valuable because they allow you to have specific experiences (*e.g.*, appreciating fine wine) that are higher up in the objective list than other experiences.

Both these rationales gesture toward some kind of value-neutral standard for comparing capabilities. Ultimately, however, the objective list theorist cannot get around the need to make naked judgments of value. As Sunstein concludes: “Those who are able to run, or to have sexual experiences, *are better off* than those who lack these capabilities....”¹¹⁹ To substantiate these type of judgments Sunstein must find a way to combat the charge of paternalism. Before turning to the modern Capabilities Approach solution to this problem, the next section provides background on the original Aristotelian solution.

ii. Aristotle’s Broad Metaphysical Solution

¹¹⁷ *Id.*

¹¹⁸ *Id.* at 17

¹¹⁹ *Id.* at 17 (emphasis added).

Aristotle's solution to the problem of paternalism was to ground his ethics on his metaphysics. Aristotle starts with the question of being. For Aristotle, being is not a binary (to be or not to be) but rather a spectrum. The spectrum runs from less being to more being, from potentiality to actuality, from the transient to the eternal. As Anthony Kronman explains, "The love of being...It is a desire to be everlasting."¹²⁰ Therefore, the primary premise of Aristotle's metaphysics is that to be closer to eternity is to have more being.

Aristotle's spectrum of being is not merely descriptive; it also establishes a hierarchy of value. The first premise of Aristotle's ethics, then, is that a thing with more being is better off than a thing with less being. That is why all things share the longing for greater being, *i.e.* for a closer connection to eternity: "For the longing to be connected to eternity is, in Aristotle's pagan metaphysics, not merely the deepest organizing ambition of all human striving. It is the source of all striving, motion, human and inhuman, in the world as a whole."¹²¹

In the second step from metaphysics to ethics, Aristotle establishes the criteria for having more being. Aristotle argues that living things achieve greater being by actualizing the unique capabilities it has *qua* the type of living thing it is. For example, "[f]or a frog, this means the active exercise of its distinctive powers, which different frogs achieve to different degrees."¹²² Likewise, humans achieve being by actualizing the unique capacities that we have as humans, that is, by actualizing our humanity: "Human beings possess certain distinctive powers that set them apart from other living things. A man who puts these powers to work in his life exists at a higher level than those who do not."¹²³ Although there are many things that set humans apart

¹²⁰ Anthony Kronman, *To Partake in the Eternal and Divine* (draft on file with author), manuscript at 2.

¹²¹ *Id.*

¹²² *Id.* at 1.

¹²³ *Id.* at 1.

(*e.g.*, being a featherless biped), Aristotle focused on one capacity in particular: “What sets man apart from other animals, according to Aristotle, is his possession of reason.”¹²⁴

With these two steps, Aristotle now has a criteria for judging the objective value of human lives. Aristotle considers three broad life choices: the life of bodily pleasure, the life of political virtue, and the life of contemplation. The most valuable life for Aristotle is that which best actualizes the uniquely human capacity for reason: “To decide whether the happiest life is one of bodily pleasure, or political virtue, or contemplation, it would thus appear that all we need to do is determine which of these lives allows for the active exercise of reason to the greatest degree.”¹²⁵ This life, in turn, maximizes our level of being and brings us closest to achieving eternity.

For the purposes of this paper, Aristotle’s substantive conclusions are not as important as his methodology. The key to Aristotle’s defense of his objective hierarchy is his underlying metaphysical theory. The question of normative value in human lives boils down to the fundamental metaphysical question: What is it to be? The upside of this strategy is that it produces a single criterion on which to judge human lives. For any two lives, Aristotle has, in theory, a clear methodology of comparing their value. The downside, however, is that it makes Aristotle’s ethics dependent on the viability of his metaphysics. As Kronman laments, the metaphysical claim that being requires a connection to eternity is now out of date: “Nothing could be further from the modern idea that human beings live meaningful lives without a link to the eternal.”¹²⁶

¹²⁴ *Id.* at 8.

¹²⁵ *Id.* at 9.

¹²⁶ *Id.* at 2.

Without his metaphysical theory, Aristotle's ethical conclusions become arbitrary and ungrounded. The next section will explore how modern objective list theories, notably the Capabilities Approach, have attempted to re-ground Aristotelian ethics.

iii. Modern Narrow Consensus-Based Solution

Modern objective list theories, such as the Capabilities Approach, have abandoned Aristotle's broad metaphysical strategy. While holding onto Aristotle's essential ethical insights, the Capabilities Approach completely alters the aims for and justifications of the theory. To the extent that this modern strategy succeeds, it provides a useful comparison in evaluating whether Sunstein's plan can hope to avoid paternalism.

First, the Capabilities Approach drastically limits the scope of the ethical theory. Aristotle's tried to specify how someone can attain maximal being, in a sense, how one can be maximally human. The focus of his deliberation was choosing between three arguable best lives: physical pleasure, political virtue, and contemplation. In contrast, the Capabilities Approach looks at the opposite end of the spectrum, specifying what it takes to be minimally human. The focus of the Capabilities Approach is choosing what capabilities without which one is less than human. In Nussbaum's words: "I am not trying to provide an account of well-being for all public purposes. I am trying to provide an account of a central group of very fundamental entitlements...."¹²⁷ This limited scope goes a long way to alleviating concerns of paternalism. Whereas Aristotelian ethics can make infinitely fine-grained distinctions between lives, the Capabilities Approach cares only about one basic distinction: are you human or not? Therefore, the Capabilities Approach does not dictate how one ought to live one's life generally or what constitutes a fulfilling life. Unless you are arguably at the verge of losing your humanity, your life choices are not within the scope of concern.

¹²⁷ Nussbaum, *supra* note 7, at 26.

Second, within this limited scope, the Capabilities Approach grounds its substantive conclusions on consensus rather than metaphysics. Nussbaum argues that even if we cannot agree on what it means to be maximally human, we can at least agree what it takes to be minimally human: “Whatever the differences we encounter, we are rarely in doubt as to whether we are dealing with a human being and when we are not.”¹²⁸ As a result, the Capabilities Approach picks out only the capabilities that fall within the “overlapping consensus among people who have very different comprehensive views, religious and secular, of the meaning and purpose of human life.”¹²⁹ If we all agree that these capabilities are essential to human-ness, there is no need for independent justification. Therefore, the Capabilities Approach avoids the need to make underlying metaphysical assumptions. As Nussbaum explains: “This conception is emphatically *not* metaphysical; that is it does not claim to derive from any source external the actual self-interpretations and self-evaluations of human beings in history.”¹³⁰ If the Capabilities Approach succeeds in creating a list that is truly consensus-based, it avoids paternalism by eliminating room for disagreement.

Furthermore, as a pre-condition for consensus, the Capabilities Approach is self-consciously vague. Even within the narrow scope of the Capabilities Approach, Nussbaum does not expect to find consensus on specific details. The claim is only that we “share at least a very general outline of...a conception [of minimal human capabilities].”¹³¹ The goal is to allow societies and individuals to differ as much as possible in their specific beliefs while still maintaining consensus on a set of core commitments. To maximize the chances of such

¹²⁸ Nussbaum, *supra* note 22, at 215.

¹²⁹ Nussbaum, *supra* note 7, at 16.

¹³⁰ Nussbaum, *supra* note 22, at 215

¹³¹ .

consensus, the Capabilities Approach keeps its list intentionally broad and general. In Nussbaum's words, "[I]t is better to be vaguely right than precisely wrong."

Lastly, the Capabilities Approach avoids paternalism by being prospective. The Capabilities Approach argues that everyone should have certain minimum capabilities, but does not tell people that what they should do with them: "Government is not directed to push citizens into acting in certain valued ways; instead, it is directed to make sure that all human beings have the necessary resources and conditions for acting in those ways."¹³² The emphasis on future choice allows the theory to be neutral with respect to different life choices, thereby avoiding paternalism. Nussbaum wants everyone to have adequate food, but does not judge the person who chooses to fast.¹³³ If the Capabilities Approach were also retrospective, then people choices would matter. For example, suppose an individual would have chosen to fast even if he had adequate food. Prospectively, we all agree that he should have access to adequate food so that he can make that choice for himself. But retrospectively, if he lacked adequate food do we say that his lack of choice made him less than human? Given that he would have lived in exactly the same manner, it is difficult to say that he is less than human in one circumstance but not the other. By focusing on future options, the Capabilities Approach avoids having to examine people's life choices at all.

iv. Objective List Theories Applied to Torts Would be Dangerously Paternalistic

Like modern objective list theories, Sunstein lacks a fundamental metaphysical foundation for his ethical theory. Consequently, he takes a modern approach of relying on informed deliberation to generate his list of compensable capabilities. Unfortunately, the

¹³² *Id.* at 225.

¹³³ *Id.*

conditions that help the Capabilities Approach to avoid paternalism are all inverted in the context of tort compensation.

First, the scope of tort compensation contains the unlimited range of potential human harms. Consequently, Sunstein's objective list will be much more expansive than that of the Capabilities Approach. Furthermore, most tort injuries are unlikely to be so debilitating as to deprive the victim of his minimal human capabilities. Similarly, most tort victims in the United States will be starting from a high baseline of capability. Therefore, Sunstein's list will likely focus on distinctions at the upper end of the spectrum of humanity.

Second, a list of compensable torts needs to be incredibly fine-grained, and consequently, cannot be consensus-based. Sunstein's list must set prices for every conceivable tort, requiring detail at the level of fingers and toes. Furthermore, to be useful to judges, the list must avoid vagueness. If the Capabilities Approach cannot find consensus on the details of a minimal human existence, it is safe to assume that Sunstein will not be able to find consensus on the entire range of human capabilities.

Finally, tort compensation is necessary retrospective. Tort victims have usually lost a capability that no amount of compensation can bring back. The tort system cannot, for example, restore the choice of whether or not to have a leg. With the permanent loss of this capacity, the victim is forever less than fully human. This raises the theoretical question, what function does tort compensation serve? If capabilities represent the intrinsic value one's humanity, it is unclear how such a loss can be monetized. And if humanity can be monetized, does that imply we can pay people to live less-than-human lives? Such a conclusion would certainly be counter to the motivations of objective list theories such as the Capabilities Approach.

In sum, Sunstein requires a detailed accounting of the monetary value of all potential human deprivations. He has no theoretical criteria for deriving this list. Furthermore, he cannot guarantee consensus on either the ordering of the hierarchy or the monetary values assigned. Without a consensus, the list simply amounts to the majority telling the minority what their lives are worth. In short, this is the very essence of paternalism.

One might respond, however, that the government makes arbitrary determinations all the time – what is so dangerous about arbitrariness in this context? The difference is that Sunstein’s proposal requires the government to make judgments of normative value. One could easily justify a tort compensation list on hedonic grounds, for example, if the list approximated the average hedonic loss for particular injuries. The list would under compensate the minority of people who suffered above-average hedonic losses. But in that case, the list is not making any judgment about the individual victim or his hedonic suffering. The list would merely represent the average hedonic experience of tort victims. Under Sunstein’s proposal, however, the list represents the government’s judgment about the objective value of every individual’s life. Specifically, it represents how much humanity that individual has lost; a loss that no amount of hedonic wellbeing can substitute for. For the individual who feels that his loss was undervalued, Sunstein cannot offer any justification but the fact that other people disagreed.

The danger of an objective list theory of tort compensation, however, extends much farther than the individual tort victims. Because the objective list values capabilities not pain and suffering, it applies to all individuals not just tort victims. The list, for example, applies to any person who is born with a physically or mentally disability. It also applies to people who in some sense chose to have a physical impairment, such as members of the deaf community to chose not to get cochlear implants, transgendered individuals who chose to alter their bodies at the expense

of reproductive capacity, and transabled people who chose to amputate. Within a hedonic framework, you can compensate one person for the pain and suffering of his amputation while recognizing that amputation may actually increase someone else's hedonic wellbeing.¹³⁴ Within an objective list framework, however, these two positions are inconsistent; if not having a leg is bad for one person it is equally bad for another. In all these cases, therefore, the government will have to make a choice: compensating tort victims necessarily disparages the worth of other lives. While the government has a legitimate role in making normative judgments, it should not exercise this power lightly.

Therefore, the real danger of Sunstein's objective list lies in its unintended "expressive externalities."¹³⁵ To create an objective compensation list, the government must judge the relative merits of the lives of every member of society. As a result, some citizens must be ranked as essentially less human than others. The very idea of such a list goes against the core liberal value of individualism. Without either the theoretical support of Aristotelian ethics nor the self-restraining safeguards of the Capabilities Approach, Sunstein's objective list theory offers no grounding to make such binding normative judgments.

Conclusion

Sunstein and U&L believe they see a problem with the hedonic theory of tort compensation. Namely they believe that it cannot support compensating victims that we intuitively want to compensate. In this paper, I have argued that authors overstate the problem. A nuanced understanding of overestimation and underestimation biases reveals that we do not need a new theory to justify compensating victims of catastrophic loss. Furthermore, even if this were

¹³⁴ See, e.g., Sabine Mueller, *Amputee Envy: People with Body Integrity Identity Disorder Feel Alienated From a Part of Their body and Want to Have it Amputated*, SCIENTIFIC AMERICAN MIND, Dec 2007, at 60.

¹³⁵ C.f. Dov Fox & Christopher L. Griffin Jr., *The Collateral Effects of Law on Social Behavior: The Case of Antidiscrimination Law and Selective Abortion*, available at <http://ssrn.com/abstract=1129905> (examining the unintended expressive externalities of the Americans with Disabilities Act).

a problem in tort law, Sunstein and U&L fail to offer attractive solutions. Both authors appeal to non-hedonic theories of wellbeing without a careful examination of the underlying philosophical claims. While these hedonic theories may apply well in other contexts, both are poorly suited to tort compensation.