

Forthcoming in *Philosophical Issues* (a supplement to *Nous*), vol. 19 on “Metaethics,” ed. Ernest Sosa, 2009.

**IN DEFENSE OF FUTURE TUESDAY INDIFFERENCE:
IDEALLY COHERENT ECCENTRICS AND THE
CONTINGENCY OF WHAT MATTERS**

Sharon Street
New York University

1. Introduction

Some strange characters inhabit the world of metaethics. I’m not talking about the people who do metaethics, of course, though no doubt we’re an odd group too. The characters I have in mind are purely hypothetical, and they’re distinguished by two main features. First, they accept some value that is utterly unheard of, morally repugnant, or both. Second, their acceptance of this value coheres perfectly, as a logical and instrumental matter, with all of their other values in combination with the non-normative facts. Call these characters *ideally coherent eccentrics*.

The ranks of ideally coherent eccentrics—or *ICEs*, as I’ll call them for short—include some well known figures. Perhaps the most famous is a character suggested by one of Hume’s remarks in the *Treatise*: this is the man who prefers the destruction of the whole world to the scratching of his finger.¹ Also very famous is Rawls’s blade-of-grass counter, whose sole pleasure is “to count blades of grass in various geometrically shaped areas such as park squares and well-trimmed lawns.”² More recent additions to the cast of ICEs include two figures described by Gibbard: an ideally coherent anorexic, who “accepts norms that prescribe death by starvation, if the alternative is a figure plump enough to sustain life,”³ and an ideally coherent Caligula, who “aims solely to maximize the suffering of others.”⁴ The final ICE I’ll mention is perhaps the most bizarre character of all, and another of the most famous. This is Parfit’s man with Future Tuesday Indifference, who, given a choice, always prefers agony on a future Tuesday to the slightest pain on any other day.⁵

Ideally coherent eccentrics are generally regarded as heavy hitters on behalf of the following view: *Some values and desires are intrinsically irrational*.⁶ On this view, there are some values and desires such that even if an agent is entirely internally consistent in holding them and is making no mistakes whatsoever about the non-normative facts, that agent is nevertheless making a *normative mistake* in the sense that he or she is valuing something that is not in fact valuable, or desiring something that is not in fact desirable. To put the view in reasons terminology,⁷ there are facts about how an agent has most normative reason to live that hold independently of that agent’s evaluative attitudes and what follows, as a logical or instrumental matter, from those attitudes in combination with the non-normative facts.⁸ Let’s call views of this sort *attitude-independent conceptions of normative reasons*.⁹ According to *attitude-dependent conceptions*, in contrast, there are *no* facts about how an agent has most normative reason to live that hold independently of that agent’s evaluative attitudes and what follows from within the standpoint constituted by them; instead, an agent’s normative reasons are always ultimately a

function of that agent's own evaluative attitudes and what is logically or instrumentally entailed by those attitudes in combination with the non-normative facts.¹⁰

In this debate—which we may regard as a modern, secular version of the *Euthyphro* debate—ICEs are commonly thought to tell strongly in favor of attitude-independent conceptions. For as it has seemed to many:

- (1) Whatever one's desires, the destruction of the world is not in fact preferable to the scratching of one's finger.
- (2) No matter what a person's tastes, counting blades of grass is an unworthy life's work.
- (3) Regardless of what one prefers, it is irrational to starve oneself to death for the sake of a trim figure.
- (4) Even if one enjoys inflicting pain on others more than anything else, one does not have most normative reason to torture people for fun.
- (5) Whatever one's evaluative attitudes, the fact that agony will occur on a future Tuesday is no good reason whatsoever to choose it over a slight pain on any other day.

These are powerful, widely shared intuitions, and many theorists regard them as providing a knockdown case against attitude-dependent conceptions of normative reasons, which (so it is said, and I agree) imply the following:

- (1') If the choice arises, Hume's man has most normative reason to choose the destruction of the world over the scratching of his finger.
- (2') Rawls's man has most normative reason to devote his life to counting blades of grass.
- (3') The ideally coherent anorexic has most normative reason to starve herself to death.
- (4') The ideally coherent Caligula has most normative reason to torture people for fun.
- (5') If you have Future Tuesday Indifference, then the fact that agony will occur on a future Tuesday is good reason for you to choose it over a slight pain on any other day.

Many people find (1')-(5') so implausible as to constitute a virtual *reductio* of any theory that implies them. When confronted with these alleged implications, attitude-dependent theorists may adopt one of two main strategies. The first strategy is to argue that contrary to superficial appearances, claims such as (1')-(5') are *not* in fact consequences of an attitude-dependent view. For example, regarding (3') and (4'), one might argue that ideal coherence rules out values such as starving oneself to death for the sake of a trim figure or torturing other people for fun.¹¹ This is an important possible strategy, and shouldn't be overlooked. I'm skeptical of its prospects, however, and for the most part set it aside in this paper.¹² My goal is to focus on the second main strategy, which is to agree upfront that (1')-(5') *are* consequences of an attitude-dependent view, and then to argue that these consequences are not at all as implausible as is usually thought. In my view, claims (1')-(5') are true and claims (1)-(5) are false, and the most plausible overall theory of value will come straight out and say so.¹³

The debate between attitude-independent and attitude-dependent conceptions must be decided in a holistic fashion. Elsewhere I argue that attitude-independent conceptions face unacceptable epistemological problems when we seek to reconcile them with our best naturalistic understanding of the world.¹⁴ In this paper, I set these arguments entirely to one side, and focus instead on the following issue. Even assuming for the sake of argument that attitude-independent conceptions face major epistemological challenges, one might think that attitude-dependent conceptions nevertheless have implications even *more* implausible, such that an attitude-independent conception is *still* the most plausible view all-things-considered.¹⁵ My aim here is to

rebut such challenges by showing that implications such as (1')-(5') are much more plausible than is commonly assumed—indeed, I'll suggest, to the point of being rather intuitive.

It is impossible to discuss all of the ICEs that resourceful attitude-independent theorists might imagine.¹⁶ Indeed, one of my main points in this paper will be that discussions of ICEs require a lot more care than they usually get; one cannot just toss out a roughly sketched ICE and after brief discussion expect to have established anything philosophically useful. The paper's strategy, therefore, is to undertake three case studies, each of which raises distinct issues. I focus on the ideally coherent anorexic (section 2), the man with Future Tuesday Indifference (section 3), and the ideally coherent Caligula (section 4). As will become clear, it is impossible to give even these three cases fully adequate treatment in one paper; my discussion of Caligula, in particular, must be limited to a sketch. The paper's aims are general, however. From each case study I draw several "lessons," the cumulative effect of which is, I hope, to lay out a general strategy for handling ICEs and to show that the plausibility of attitude-dependent conceptions of normative reasons is routinely seriously underestimated.¹⁷

2. First case study: The ideally coherent anorexic

As Gibbard describes her, the ideally coherent anorexic "accepts norms that prescribe death by starvation, if the alternative is a figure plump enough to sustain life," and she is stipulated to be entirely coherent in doing so, such that she is making no mistakes whatsoever about the non-normative facts and "the norms she accepts tell her to starve, and the higher order norms she accepts tell her to accept those lower order norms."¹⁸ According to Gibbard, while the anorexic may view herself as acting rationally, "the point here is, the rest of us do not. We think it irrational to starve oneself to death for the sake of a trim figure" and that this is so no matter what set of norms we or the anorexic or indeed anyone else accepts.¹⁹ In Gibbard's view, the anorexic is failing to recognize an attitude-independent fact about how she has most normative reason to live (where this claim about attitude-independent normative facts is understood as a substantive normative one, internal to normative discourse²⁰).

Gibbard thinks most readers will share his intuition about the case. Relying on this, he takes it to be established that a metaethical view should make it possible for us to go on making these kinds of extremely strong claims about attitude-independent normative reasons; the intuition about the ideally coherent anorexic's reasons is accepted as a fixed point that a satisfactory metaethical view should preserve our "right" to affirm.²¹ Though it is not obvious at the time, this is a pivotal moment in Gibbard's overall case for expressivism and quasi-realism in *Wise Choices, Apt Feelings*. For it is Gibbard's hope of accommodating such extremely strong claims of "validity independent of acceptance"²² that motivates Gibbard to seek an expressivist and quasi-realist interpretation of them—rather than, for example, proposing a subjectivist reforming definition that construes normative claims as straightforwardly truth-apt. The point here isn't to get into assessing expressivism and quasi-realism,²³ but rather to draw attention to the philosophical weight that is being placed on the ideally coherent anorexic's shoulders, and to suggest that if she is going to be asked to bear this much of a load, we need to look at her with more care.

To begin with, as is obvious but worth drawing explicit attention to, anorexia nervosa is a serious, real-life illness that is rampant among women in our culture.²⁴ Appeal to this example therefore risks trading on our intuitions not about the case at hand, which is purely hypothetical, but rather about real-life women. This doesn't mean we can't use the example, but it does mean that we need to move slowly. There is substantial philosophical risk that when we ask ourselves "Would an ideally coherent anorexic have most normative reason to starve herself to death for the sake of a trim figure?" the character we unthinkingly imagine is much like the women we know

from real life, with just a tweak or two. We may even think vaguely of anorexics we have known. And naturally, if we are even implicitly calling to mind the sorts of women we know from real life, we will feel strongly that none of them could ever have normative reason to starve herself to death for the sake of a trim figure; indeed, we bristle at the suggestion, which actually seems quite offensive.

None of this, however, tells the least bit against an attitude-dependent conception of normative reasons. For—in a point that is also obvious but often overlooked in practice—if one is going to appeal to intuitions about cases to help decide between attitude-dependent and attitude-independent conceptions of normative reasons, then one is under a strict philosophical obligation to imagine the cases in question accurately; otherwise one’s intuitions concerning them are of no relevance to the dispute at hand.

So let’s now spend some time making sure we’re imagining the case accurately. I’m going to argue for two main points. First, no real-life human being could ever be an ideally coherent anorexic; attitude-dependent conceptions of normative reasons therefore have nothing to fear from our intuitions about real-life people. Second, if we take the time to imagine clearly what a genuinely ideally coherent anorexic would look like, it becomes intuitively plausible that she would have most normative reason to starve herself to death for the sake of a trim figure; attitude-dependent conceptions of normative reasons therefore do a good job of capturing our intuitions about these cases too.

Start with real-life human beings. Any real-life person has multiple, deeply held ends, many of which require being alive to promote or pursue. Any real-life woman,²⁵ for example, presumably cares about the lives of her family and friends; she has a career or another broad plan of life that she hopes to pursue; she may be interested in travel or science or how the Braves are going to do next season; and so on. Even if she is suffering from severe anxiety or depression, she will be able to agree, if we ask her, that she at least used to care passionately about such things, and would do anything if she could just feel like her old self again. Such utterly typical human goals and values, whatever their exact nature, coupled with the fact that being alive and not dead due to starvation is usually necessary to promote them, ensure that no real-life woman will ever have most normative reason to starve herself to death for the sake of a trim figure—and this even according to an attitude-dependent conception of normative reasons.

One might object to this as follows. If a real-life woman is suffering from anorexia, then she will presumably care very strongly about being thin. And couldn’t there be at least some real-life cases in which this desire is so strong that it overpowers all of the woman’s other desires—to the point where starving herself to death is indeed what she has most normative reason to do, according to an attitude-dependent conception of normative reasons?

To think this, however, is to fail to think carefully enough about real-life women and what would be required to be an ideally coherent anorexic. First, granting for the sake of argument that any real-life woman suffering from anorexia wants—perhaps even very strongly—to be thin, the figure she actually wants is, contrary to what Gibbard imagines in his example, “plump enough to sustain life.” After all, the figure that any real-life woman suffering from anorexia presumably wants is just a supermodel figure, and a supermodel figure is perfectly capable of sustaining life, as supermodels themselves demonstrate. The real problem is that given the average woman’s build (not to mention constraints of time and money), a supermodel figure is impossible to achieve no matter what she does. But then it’s not as though the real-life anorexic is in the following coherent state of mind: she really wants a supermodel figure; she is under no illusions about the fact that she cannot attain this figure no matter what she does; nevertheless she values the figure in question so highly—indeed so much more highly than any other end that would require her to go on living—that what she wants more than anything is to die

in what she herself regards as a doomed pursuit of this figure. But this is exactly the kind of state an ideally coherent anorexic would have to be in.

As anyone familiar with real-life anorexia knows, moreover, it is not as though what drives the condition is love of a “trim” figure as an end in itself—as if the victim viewed having a thin, supermodel body as an intrinsically valuable state of being. What drives the real-life condition are much more familiar human desires such as the desire to be loved, to be thought beautiful and desirable, to feel in control of one’s life, and so on—all of this coupled with the false and perhaps largely unconscious assumption, fostered by a particular cultural setting, that a certain type of body is a precondition of all this. As anyone familiar with real-life anorexia also knows, the victim often suffers from grossly inaccurate perceptions of her own body. While to friends and family she looks emaciated and sick, the victim herself looks in the mirror and sees someone who is “fat.” She believes she is overweight compared to the models in magazines; she believes other people look at her arms and think they are flabby, and so on. Friends and family who dispute these assumptions, she falsely believes, are lying or just trying to be nice. All of this is painfully far from ideal coherence.²⁶

If a person has even remotely statistically normal human traits, then, an attitude-dependent conception of normative reasons will not imply that she has normative reason to starve herself to death for the sake of a “trim” figure. On the contrary, the attitude-dependent conception will imply that any real-life woman who ever thought this would be making a terrible mistake on her own terms. Attitude-dependent conceptions of normative reasons therefore have nothing to fear when tested against our intuitions about the normative reasons of real-life women suffering from anorexia.

Still, what about the hypothetical case? Even if the ideally coherent anorexic will never exist in real life, it might seem deeply implausible to claim that were such a person ever to exist, she would have most normative reason to starve herself to death. Yet this is exactly what an attitude-dependent conception of normative reasons implies (or at least so I wish to grant upfront). Why isn’t this a major point against attitude-dependent conceptions?

Before continuing, let me comment briefly on a methodological point. Some might think that it is silly, irrelevant, or in some other way misguided to test conceptions of normative reasons against our intuitions about far-fetched, imaginary cases. Fellow defenders of attitude-dependent conceptions in particular might think that it is a mistake even to agree to talk about such cases. This is not my view. I agree with those who think that the best overall theory of value must prove itself in its handling of hypothetical cases. Indeed, I would go farther than this: not only is it not silly to think about ideally coherent eccentrics, these characters haven’t gotten anywhere near *enough* philosophical attention.

In my view, ICEs are where the action is if we want to get clear on the relation between our attitudes, value, and the world. It is impossible to make a full case for this point here, for the case depends, among other things, on the claim that attitude-dependent theorists can satisfactorily explain the alleged “values” of logical and instrumental consistency. Assuming that they can explain this, though—and I think they can²⁷—all eyes should turn to ICEs. For as the forgoing discussion helps to bring out, attitude-dependent and attitude-independent conceptions of normative reasons will tend to generate many of the same, intuitively plausible consequences about the reasons of real-life human beings. It is only in cases where ideal coherence is stipulated that we can be sure the two views will generate different consequences; anything less and the attitude-dependent theorist may be able to account for what is intuitively an error in an agent’s judgments about his normative reasons by pointing out inconsistencies or factual or instrumental mistakes. This last comment might prompt the question: If it’s only in bizarre hypothetical cases that attitude-dependent and attitude-independent conceptions are sure to generate different results,

then why does this debate matter at all? If we will never face these kinds of agents in real life, in other words, what does it matter if we disagree about what to say about them? I think it matters tremendously in a way that has to do with our conception of ourselves and our place in the world. I say a little more about this in the concluding section.

Turning, then, to the purely hypothetical case of the ideally coherent anorexic, what I hope is already becoming clear is that in order to be an ideally coherent anorexic, a person would have to be an alien being—if not literally, then something very close to it. This is because in order to be an ideally coherent anorexic, a person would have to value starving herself to death for the sake of a “trim” figure more than anything else in the world—more than anything else that required continuing to live, anyway—and she would have to be free from all deeper normative commitments according to which this is a stupid and unworthy value. Her commitment to this value would have to be consistent with all non-normative facts about her situation, including facts about how she acquired this value in the first place and facts about alternative sorts of lives that it would be open to her to lead. She would need to love the figure in question more than life itself, knowing without a shred of illusion that it wasn’t compatible with survival, but nevertheless valuing it with such utter devotion that she was willing to sacrifice everything in order either to achieve it just for a brief moment or else to die in its pursuit.

If one forces oneself to imagine what such a being would actually have to look like, then I think one will either cease to have clear intuitions about the case, or else conclude that someone so clearheaded and impassioned in the pursuit of her goal might well have most normative reason to starve herself to death. “Go for it,” we might say to her, after long talks with her in which we assured ourselves of her flawless consistency and lack of illusions about the non-normative facts. “I guess you should do whatever floats your boat, if you really feel that strongly about it.” Our attitude here might be much like the attitude many of us would take toward a person who knowingly accepts a tremendous risk of dying in order to have a shot at reaching the summit of K2. The ideally coherent anorexic’s aesthetic is harder to relate to, no doubt, but it doesn’t seem to be in a fundamentally different category. On the contrary, we can imagine the ideally coherent anorexic thrilled by those brief moments of achieving the figure she regards as so beautiful in the same way that a climber might be thrilled by his brief moments on the summit of K2, even while knowing full well that the gathering storm means he won’t make it back down alive.

Putting aside moral considerations such as a concern about friends and family these characters might be leaving behind—I address moral considerations separately in section 4—my own intuitive reaction to the case is that the ideally coherent anorexic has most normative reason to starve herself to death in the attempt to achieve this figure that she values with every fiber in her body, just as I think the occasional real-life human being could have most normative reason to die in the attempt to ascend K2. At the very least, I hope it is clear that any intuitions one might retain to the contrary are not at all strong enough to support the argumentative weight that Gibbard puts on them. More robust intuitions need to be located before we have adequately motivated the idea that a satisfactory metaethical view must be able to accommodate strong attitude-independence claims such as this claim about the ideally coherent anorexic.

From our consideration of this case, we may draw three more general lessons. They are obvious, but worth stating explicitly because of how easy they are to overlook in practice.

Lesson 1: When discussing ICEs, one is under a strict obligation to make sure one is imagining these characters in full and accurate depth, otherwise one’s intuitions regarding them are of no philosophical relevance.

Lesson 2: Vigilance is required to avoid the pitfall of assuming that ICEs look pretty much like ordinary human beings with just a tweak or two. In many cases, an accurately imagined ICE will look more like an interesting visitor from another planet than a human being.

This “alien visitor” analogy is worth keeping in mind, since one’s intuitions about the reasons of alien visitors are likely to be more plastic than one’s intuitions about a character one is (unconsciously and perhaps mistakenly) assuming to be a human being.

Lesson 3: Care is also required to avoid underestimating the resources of an attitude-dependent conception of normative reasons to yield the intuitive result about the reasons of real-life human beings. Real-life human beings are never ideally coherent, and appeals to mistakes about the non-normative facts plus inconsistencies with a person’s most deeply held values can often get one much farther than one might think. There are deep similarities in what real-life human beings tend to care about and the circumstances in which they tend to find themselves, and these similarities are often enough to underwrite most of normative commonsense.

3. Second case study: The man with Future Tuesday Indifference

As our second case study, let us turn to a man who makes the ideally coherent anorexic look almost normal. This is Parfit’s man with Future Tuesday Indifference, whom Parfit describes this way:

Consider...an imagined man who has an attitude that we can call *Future Tuesday Indifference*. This man cares about his own future pleasures or pains, except when they will come on any future Tuesday. This strange attitude does not depend on ignorance or false beliefs. Pain on Tuesdays, this man knows, would be just as painful, and just as much *his* pain, and Tuesdays are just like other days of the week. Even so, given the choice, this man would now prefer agony on any future Tuesday to slight pain on any other future day.²⁸

Appealing to our intuitions, Parfit then writes:

That some ordeal would be much more painful is a strong reason *not* to prefer it. That this ordeal would be on a future Tuesday is *no* reason to prefer it. So this man’s preferences are strongly contrary to reason, and irrational.²⁹

This case and others like it, Parfit thinks, provide some of the clearest examples of intrinsically irrational preferences,³⁰ and Parfit rests a great deal of his case for what he calls an “objectivist” theory of reasons on our intuitions about them.³¹

And indeed these intuitions seem strong. For the sake of discussion, it will be useful to imagine a concrete example of a person with Future Tuesday Indifference. Let’s call our man *Indy*, and imagine the following exchange between Indy and his dentist:

Dentist: Indy, I’m afraid that molar needs to come out. Let’s see, I have two openings next week, one on Tuesday, the other Wednesday. The thing about the Tuesday appointment is that I’m afraid we’re currently out of anesthetic, and we won’t get our next shipment until Wednesday morning. So if you come in Tuesday, we’ll have to do the procedure without anesthetic. On Wednesday, though, we can do it with anesthetic in the usual way. Am I right in assuming you’d prefer to come in Wednesday?

Indy: Hmm, I’m free both days, but please schedule me for Tuesday.

Dentist: Are you sure? If you come in Wednesday, all you’ll feel is a quick pinch when I inject the anesthetic, and nothing unpleasant after that. If you come in Tuesday, though, then—I hate to say it—but it will be agony.

Indy: Sure, sure, I understand. No, please book me for Tuesday. I know it’ll be agony, but the thing is, it’ll be on a Tuesday. I’d much prefer to undergo that agony on Tuesday than to undergo the quick pinch on Wednesday.

No doubt it sounds like Indy is making a serious mistake about his normative reasons. As I'll now try to show, however, this initial response to the example depends on a failure to think carefully enough about this complicated case.

Before going further, it is essential that we clarify how pain is being understood. I will work with Parfit's own understanding of pain, which seems right to me in general outline.³² On Parfit's view, the experience of pain involves a sensation coupled with what Parfit calls a "hedonic disliking" of that sensation.³³ This hedonic disliking, Parfit says, is not intrinsically rational or irrational. For example, if someone for some reason has no hedonic dislike of the sensation of burns to his skin, perhaps because of a medication he is on or an unusual physical condition, then that lack of hedonic dislike is not irrational, any more than failing to dislike the sound of nails on a chalkboard would be irrational. Rather, Parfit says, we should see the hedonic disliking of a sensation as what makes that sensation painful, and what confers badness upon it; in the absence of our hedonic dislike of a sensation while having it, Parfit thinks, that sensation isn't painful or in and of itself anything we have normative reason to want to avoid.

The experience of pain, then, on Parfit's understanding, is not just a simple sensation but the structurally more complex conscious state of having a sensation that one hedonically dislikes. The more intense the hedonic dislike, the more intense the pain. This allows us to fill out Indy's case in important additional detail. In particular, if Indy has his tooth removed on Tuesday, without the anesthetic, he will experience sensations—including, for example, the sensation of his molar being wrenched out of his gum—that will be accompanied by extraordinarily high levels of hedonic dislike; this is just what it is to be in agony. In contrast, if Indy has his tooth removed on Wednesday, he will experience the sensation of the anesthetic injection into his gum, coupled with a mild hedonic dislike of the sensation, but that hedonic dislike will be nothing at all in intensity compared to the hedonic dislike he will experience of his sensations if he does the extraction on Tuesday.

I won't dwell on the point here, but I believe that in accepting this account of pain, Parfit makes an important, though not decisive, concession to attitude-dependent conceptions of value. For on this account of pain, when we dislike some present sensation, it is this dislike that makes our conscious state bad. Hedonic likings and dislikings, Parfit is ready to agree, can be claimed to create or confer value or disvalue.³⁴ This is an important concession because Parfit is explicitly acknowledging that what attitude-dependent theorists argue is *always* ultimately going on in cases of value does at least *sometimes* go on, namely that value or disvalue is conferred upon the world (in this case, upon our own sensations) by evaluative attitudes which in and of themselves are ultimately neither rational nor irrational. Parfit is thus granting that the dependence relation between value and our evaluative attitudes at least *sometimes* runs in the direction claimed by those of us who favor the attitude-dependent answer to the Euthyphro question. The general idea of an attitude's conferring value on the world at least makes sense.

Again, though, I won't dwell on this, for Parfit thinks hedonic likings and dislikings are a special case. Recall that pain is being understood as a complex state of a sensation plus a hedonic disliking of it. Parfit points out that we can have higher-level desires directed at this complex state, and he calls these attitudes *meta-hedonic desires*. For example, any statistically normal human being will have a strong meta-hedonic desire not to experience the combination of sensations and intense hedonic dislike involved in having his or her tooth extracted without anesthetic. According to Parfit, there are crucial differences between hedonic likes and dislikes and meta-hedonic desires. For example, whereas hedonic dislikings of the kind involved in pain are present only at the time of the sensation involved in the pain, a meta-hedonic desire needn't occur at the same time as the pain experience. And whereas hedonic likings and dislikings aren't intrinsically rational or irrational, meta-hedonic desires can be.

This is where Indy comes back in. For as we may now put it, Indy has a meta-hedonic preference to suffer agony on Tuesday rather than undergoing a mild pain on Wednesday. And Parfit's claim—still extremely plausible for all I've said so far—is that this *meta-hedonic* preference is the one that's intrinsically irrational. This, in other words, is where Parfit wants to make his stand for an attitude-independent conception of normative reasons. The claim is that some things—our own pain, for example—supply us with normative reasons in a way that is robustly independent of our *meta-hedonic desires*. This is a useful way to clarify what is at issue between attitude-dependent and attitude-independent conceptions of normative reasons, and I accept it here for the sake of argument, even though I think Parfit is wrong to suggest that there is a bright line between hedonic likes and dislikes and meta-hedonic desires. The point I'm concerned to argue for is this. While it's less obvious because of the complexity of the attitudes involved, the dependence relation between value in general (for example, the disvalue of pain) and meta-hedonic desires is ultimately best understood on the same model as the dependence relation that Parfit has already admitted exists between the value of certain bodily sensations (for example, hearing the sound of nails on a chalk board) and hedonic likes and dislikes.

To see this, let's now look more closely at Indy, forcing ourselves to imagine in vivid depth the kind of thing that would be involved in genuinely possessing a meta-hedonic preference for pain on future Tuesdays. First let's imagine what it's like to be Indy as he contemplates the two appointment options for next week. It must be something like this. When Indy considers the agony of the procedure with no anesthetic, he is perhaps briefly horrified like the rest of us, imagining the sensations plus the intense hedonic disliking of them. But as soon as he hears that it would be on a Tuesday, he sighs in relief, feeling a surge of relaxation and peaceful unconcern. He doesn't know what it is about Tuesdays, but for some reason he isn't bothered at all by the idea of agony on that day of the week. He feels this way even when he imagines as vividly as possible the intense hedonic dislike he knows he'll feel toward the sensations of his tooth being wrenched out. "It's odd," he says, when you ask him to say more. "It's almost as though the more agony I know I'm going to experience—in other words, the more intensely I know I'm going to dislike the sensations—the more calm I feel in contemplating it—again, so long as I know it's Tuesday we're talking about—that part is crucial. When I contemplate the prospect of pain on any other day of the week, it's completely different—I'd just *really* rather not go through it, even if I know it's going to be mild."

That's Indy at the time he is making the appointment. What is it like to be Indy on Tuesday when he goes in for the extraction? Here we reach a fork in the road, for there are two ways to imagine the case. We know that looking ahead to pain on a *future* Tuesday, Indy has a meta-hedonic preference for agony on that day rather than mild pain on other days. But what happens when Tuesday arrives? Does Indy's meta-hedonic indifference to pain on that day persist throughout the day, or does it disappear at 12:00 a.m. Tuesday morning, his meta-hedonic preferences suddenly reverting to normal human ones the moment the Tuesday in question switches from being future to being present? Parfit's description in *Reasons and Persons* makes it clear that he is thinking about the case in the latter way.³⁵

We will need to consider both versions of the case, however. The reason for this is that the Future Tuesday Indifference example plays upon intuitions from many sources at once. In particular, as I'll try to bring out, the example gains its seeming force by appealing simultaneously to intuitions having to do with the nature of pain, the arbitrariness of Tuesdays, personal identity, prudence requirements, and indeed morality itself. To respond, the attitude-dependent theorist needs to tease apart these different elements as much as possible and address them in turn. To that end, let us distinguish the two types of Future Tuesday Indifference as if we were adding new entries to the DSM-V. Call the condition in which the meta-hedonic preference for pain on Tuesdays is consistent over time *Consistent Tuesday Indifference (CTI)*, and let us

suppose that this is Indy's condition. Let us call the other condition, in which the preference regarding pain on Tuesdays reverts to normal as soon as a future Tuesday becomes present, *Future Tuesday Indifference with Present Tuesday Horror (FTI-PTH)*. This label calls attention to one of the latter condition's distinguishing features, namely that every present Tuesday a person who has this condition is horrified by what his past self has set him up for.

Consistent Tuesday Indifference is the simpler case and I wish to start with it; doing so will enable us to focus our attention on intuitions about pain and the arbitrariness of Tuesdays before addressing the more complicated case. Assuming that Indy has CTI rather than FTI-PTH, what is it like to be Indy when he goes in on Tuesday for the extraction? We can imagine it's something like this. All morning Indy goes about his usual business, giving no more thought to his afternoon dental appointment than he would to an afternoon haircut. He arrives at the appointment on time and settles into the dentist's chair, nonchalant and at ease, his mind focused on the book club meeting he is hosting that night—wondering whether he should stop off on the way home to buy more chips. Then the dentist begins the procedure. The intense hedonic dislike kicks in. This is *unbelievably* painful. Indy writhes around in the chair, moaning involuntarily, and breaking into sobs on and off. He experiences searing feelings of hatred and resistance to these sensations; it is as though his physical self is in all out revolt, begging the dentist to stop. This, we may imagine, is his intense hedonic dislike in action.

Strangely enough, however, at the same time he's experiencing all this, Indy's core emotional being remains in a state of deep calm. Perhaps there's a sense of extreme focus and concentration on what his body is undergoing that makes Indy feel—at the level of his innermost self—more curiosity and detached interest than anything else as he contemplates these terrible sensations and his hedonic hatred of them. When we ask him to tell us more about the experience, he invokes an analogy often used by students and teachers of meditation, comparing himself to a lake whose deep waters remain still even as its surface is whipped by violent waves during a storm.³⁶ Indy hastens to add that he doesn't practice meditation and is no Buddhist—he's just always felt this way about pain on Tuesdays, never having trained for it—but he suspects that his experience on Tuesdays is at least a bit like the inner experience of monks whose training has enabled them to achieve such concentrated emotional distance from horrendous bodily pain that they are able to sit cross-legged and perfectly still as they burn to death in an act of self-immolation.³⁷ We can imagine that when the agony of the tooth extraction subsides, Indy calmly wipes his eyes and doesn't give the episode a further thought, his meta-hedonic indifference to pain on Tuesdays manifesting itself in the days and months that follow with serene, untroubled memories of the Tuesday tooth extraction. All of this, we may suppose, is Indy's meta-hedonic indifference to pain on Tuesdays in action.

Had Indy scheduled the appointment for Wednesday, we can imagine, it would have been at least a small source of irritation or distraction all week ahead of time until he had it; no analogous sense of calm would have accompanied the experience before, during, or after. He might explain to us that when he contemplates the possibility of a mildly painful injection on Wednesday (or any other day that isn't Tuesday), feelings of annoyance, frustration, and anxiety well up inside him. "I know it'll just be a quick pinch accompanied by a mild dislike of that sensation," he might say, "but I'm telling you, the thought of having to go through that on a non-Tuesday I just find aggravating." This, we may suppose, is the flipside of Indy's meta-hedonic preference for pain on Tuesdays in action.

To bring the case intuitively home, let's finish by imagining briefly how Indy might have ended up with this strange motivational structure. This genealogical story, I want to emphasize, is completely irrelevant from a normative point of view: Indy's normative reasons are what they are no matter how he got to be the way he is, and strictly speaking it wouldn't matter if he had popped out of a hat with this motivational structure. Above all, I am not suggesting that the

particular origin of his preference justifies it in any way; this is false. My reason for sketching a causal story about Indy's origins is purely heuristic—the point is simply to aid the imagination. Because the case involves such a bizarre configuration of preferences, it is extremely difficult to achieve a clear intuitive grasp of the relevant motivational structure without giving some thought to how a creature could have ended up with it. In other words, we aren't being fair to the attitude-dependent theorist unless we test our intuitions against an *accurate* picture of Indy's motivational structure, and I think it's almost impossible to get such a picture in one's head without at least briefly imagining a story along the lines I'm about to offer.

First of all, as I hope is obvious, Indy is best imagined as an interesting visitor from another planet; members of *homo sapiens* simply aren't wired like this. Even if some Buddhist monks might achieve a similar meta-hedonic detachment from their own agony, their condition still bears little resemblance to Indy's in that their "indifference" to their own hedonic experience is something achieved and sustained only by an extreme effort of concentration, and it is also obviously not something that comes and goes with Tuesdays. Thus, while Indy might live and work on planet Earth, and while he might look and act a lot like the rest of us apart from the fact that he seeks to schedule all remotely painful experiences for Tuesdays, he is clearly not one of us.

Instead, let's imagine, Indy is a member of a species that evolved on a planet in a nearby solar system. The planet is a lot like Earth, but with a few exceptions. In particular, we might imagine, for millennia until recently, Indy's species evolved right alongside another one, a cruel bird of prey that spent most of its time asleep but periodically emerged to feed, mate, and practice its hunting skills by tormenting members of Indy's species. We can imagine that there were no hiding places available, and that horrendous physical torture and pain were the norm whenever the birds were awake and active. We can also imagine that the birds slept and woke up as a group, their activity cycle coordinated with the cycle of one of the planet's moons. In this evolutionary environment, we can imagine, it promoted survival and reproduction for members of Indy's species to be emotionally hardwired so as not to worry about—and indeed to be meta-hedonically indifferent to—pain inflicted during the birds' waking periods; suffering these excruciating attacks was so regular and inevitable that it paid (from the standpoint of reproductive success) not to give the attacks the slightest thought in advance, and indeed to retain a deep sense of emotional calm when they occurred, in spite of the hedonic agony. When the birds were dormant, in contrast, it paid to have the meta-hedonic desires regarding pain of the kind we're all familiar with: pain (and the bodily injury it signals) was avoidable during those periods, so it increased reproductive success to fret about it in advance and hate it at a meta-hedonic level when it occurred.

Something along these lines, we may imagine, is Indy's evolutionary heritage, and it explains why his entire emotional being is organized so as to be meta-hedonically indifferent to pain that falls during certain regularly occurring periods. Indy may or may not know about his evolutionary origins; it doesn't matter. The important thing is that he feels total meta-hedonic indifference to pains that fall during certain regularly occurring intervals of time. And on Earth, where we may suppose Indy was born and continues to live, these regularly occurring intervals of time are Tuesdays.³⁸

We have now a more developed picture of what a man with Consistent Tuesday Indifference might look like, and he has almost all of the features Parfit imagines in his original example—with the important exception of Present Tuesday Horror, which we'll turn to in a moment. First, Indy's pains are unquestionably pains: they consist in certain sensations coupled with intense hedonic dislike of them.³⁹ Second, he has a meta-hedonic preference for experiencing even agony on Tuesdays as opposed to mild pains on any other day. Third, this meta-hedonic preference is not based on false beliefs or ignorance: Indy knows full well that the

pain he will experience on Tuesday is much more severe than the pain he would experience on Wednesday, but he is nevertheless indifferent to it at the meta-hedonic level, in a way that coheres perfectly with all of his other meta-hedonic desires. As a result of his evolutionary heritage—which, as I wish to emphasize again, has no *justificatory* relevance whatsoever, and is mentioned just to help us imagine vividly his motivational setup—it is built into his emotional structure at the deepest level to prefer, other things being equal, pain on Tuesdays to pain on any other day. This is so in just the same way as it's built into our own emotional structure at the deepest level to prefer, other things being equal, living rather than dying, our children's happiness over their misery, to be loved and admired by others rather than hated and scorned by them, and so on.

The test now is your intuitions. I hope that this portrait of Indy helps to make it intuitively plausible that a meta-hedonic preference for pain on Tuesdays is not *intrinsically* irrational any more than a like or dislike of the sound of nails on a chalkboard is intrinsically irrational. In other words, if Indy has normative reason to try to rid himself of this preference, then the reason in question is one supplied by other preferences he might have. For example, Indy might prefer to live a long life, and it's possible that on Earth Indy's CTI causes him to take unnecessary risks on Tuesdays that lead to bodily injuries likely to shorten his life. But *this* kind of irrationality is easily accommodated by attitude-dependent conceptions of value, for it is merely a kind of instrumental irrationality relative to Indy's other preferences. What I hope has evaporated is the key intuition (or one of them, anyway) that Parfit is relying on—namely that this meta-hedonic preference is irrational *in and of itself*, quite apart from the way it coheres or fails to cohere with Indy's other preferences. From this consideration of Indy, we may draw two further general lessons about ICEs.

Lesson 4: Because of the nature of pain, examples involving pain require especially careful consideration. Pain is not just some ordinary object of our evaluative attitudes, but rather a phenomenon which by its very nature seems to involve evaluative attitudes directed at certain bodily sensations. This complicates attempts to see whether an attitude-dependent or an attitude-independent conception provides the best account of the reason-giving status of pain.⁴⁰

Lesson 5: Examples that appeal to some feature that from our point of view seems totally arbitrary—for example, Tuesdays versus Wednesdays, a year from now versus a year and a day, a mile away versus a mile and a foot, and so on⁴¹—require major efforts of imagination if one is to achieve an accurate intuitive picture of the stipulated motivational structure. Here a fanciful evolutionary story can be of use purely as a heuristic device. Human beings are organized so as to take considerations having to do with familial relations, each other's suffering, food, water, broken bones, a long life, and so on, as significant. But in principle a valuing creature could be systematically organized so as to experience anything at all as what matters above all else. Before leaping to the conclusion that such a creature would be mistaken, it is important to make sure that all instrumental errors and mistakes about the non-normative facts have been eliminated. Then one must think carefully about the creature's deep motivational structure, imagining how another creature could—merely as a result of a causal history different than our own—care just as deeply, and with just as much emotional complexity, investment, and force, about the difference between (for example) a year versus a year and a day as we ourselves care about the differences between kin and non-kin, a fractured bone and a non-fractured one, and so on.

Let us now turn more briefly to FTI-PTH, which takes everything strange about Indy and throws one more oddity into the mix. Call our person with FTI-PTH *Hortense*, to distinguish her from Indy. Though both characters have a form of Future Tuesday Indifference, it is a mistake to assume that Hortense's decisions will bear any close resemblance to Indy's. Suppose, for example, that Hortense also needs a tooth extraction and is offered the same appointment options

as Indy. As we know, what distinguishes Hortense from Indy is that although she too presently prefers the Tuesday appointment, when Tuesday at 12:00 a.m. arrives she will start caring, as Parfit puts it, “in the normal way” about pain on that day.⁴² What it’s important not to forget is that Hortense, as an ICE, *knows* this about herself. But if you care about pain in the normal way, you will be appalled by the idea of a tooth extraction without anesthetic and will refuse to go through with it if you know you could just as easily do it tomorrow *with* anesthetic. Hortense therefore knows that her Tuesday self will refuse to undergo the extraction. There are two ways she might take that fact into account. One option is not to bother making the Tuesday appointment at all, knowing that her horrified Tuesday self will just cancel it when the day arrives. The other option is to take steps now to ensure that her horrified Tuesday self will go through with the extraction. Since Parfit’s case involves the character’s actually going ahead and choosing the future Tuesday operation, we must assume that Hortense is going the second route of scheduling the appointment and planning effective steps to ensure her Tuesday self will go through with it. Anything less would be instrumentally irrational—a case of choosing what she knows to be an ineffective means to the end of having her tooth extracted—and so incompatible with Hortense’s stipulated status as an ICE.

Since Hortense knows her Tuesday self will be extremely opposed to undergoing the extraction, she will have to take strong measures to ensure compliance with her present desires. Perhaps she can hire a band of thugs to see to it that her Tuesday self is carried kicking and screaming to the appointment. The situation is complicated, though. Since it’s her own future self she is plotting against, she must take into account that her Tuesday self will know every detail of whatever plan she develops. Perhaps she can handle this by hiring resourceful thugs and directing them to plan a surprise ambush. Future Tuesday Hortense will know this is coming, however, and may hole up in a guarded place. So Hortense’s safest bet is probably to deliver herself to the thugs shortly *before* Tuesday arrives, like Ulysses lashing himself to the mast. She will need to make sure the thugs are strong enough to hold her back when 12:00 a.m. arrives, giving them strict instructions to force the tooth extraction no matter how much she begs for mercy and no matter how wrenching her meta-hedonic distress. Indeed, Hortense will need to take such measures against her future Tuesday self *whenever* she chooses agony on a Tuesday over mild pain on any other day. It may therefore make the most sense to institute a permanent regime against her Tuesday self, in which every Monday evening she commits herself to the custody of thugs who then ensure that her Tuesday self undergoes whatever agony her past self has scheduled for that day. The Tuesday self is not without her own resources, of course. She will do her best to escape, to call the police, to sweet talk the thugs into releasing her, and so on. When planning for future Tuesdays, Hortense will have to anticipate such measures and preempt them.

The picture of someone with FTI-PTH that emerges, then, is a picture of a person at war with herself: a case of captor and captive, torturer and victim, living in the same body and waging a constant battle to outwit one another—a battle made especially complicated by the fact that they have all the same information. We now see that Parfit’s description of the Future Tuesday Indifference case is misleading in a very important respect. In setting up the case, Parfit stipulates that the person has no “false beliefs about personal identity,” commenting that the man with Future Tuesday Indifference “agrees that it will be just as much him who will be suffering on Tuesday.”⁴³ We need not take any position on the nature of personal identity to observe that in the absence of further elaboration, it is natural for readers of the example to assume that the character regards his future Tuesday self as “just as much him” in some ordinary intuitive sense; it is natural for the reader to assume, in other words—perhaps without being conscious of the assumption—that the character regards his Tuesday self as “just as much him” in something roughly similar to the way in which the rest of us regard our Tuesday selves as “just as much us.” But as we’ve just seen, Present Hortense doesn’t regard Future Tuesday Hortense as “just as

much her” in anything remotely like the way ordinary people do. On the contrary, she plots against Tuesday Hortense deliberately and without mercy, seeking to predict and preempt her moves, holding her captive with a band of thugs, displaying not the slightest concern for her suffering, and so forth.

We can sidestep questions about the nature of personal identity by granting for the sake of argument that Present Hortense and Future Tuesday Hortense are one and the same person according to the “true account of personal identity,” whatever that might be, and that Hortense is perfectly aware of this fact. We can thereby respect Parfit’s stipulation that Hortense has no false beliefs about personal identity. The important thing is to diagnose a major source of the case’s *prima facie* intuitive force. In particular, if you’re implicitly assuming that Present Hortense regards Future Tuesday Hortense as “just as much her” in anything remotely like the way the rest of us do, then of course it is going to sound wildly irrational to be utterly indifferent to that self’s pain. Regular people feel at least some compassion toward their future selves and take at least some interest in what happens to them, so they are making a mistake on their own terms if they fail to look out for those later stages. But once one sees that Present Hortense regards Tuesday Hortense as “just as much her” only in a sense of the expression that is loose enough to allow the possibility of the two stages of Hortense to be mortal enemies, one may still think that Present Hortense is making an error in preferring agony for Tuesday Hortense, but now it seems to be more of a *moral* error than anything.

The most natural first reaction to the Future Tuesday Indifference case is to think that whatever else is going on, a person with this condition is making some bizarre kind of *prudential* error. But when we think carefully about what, in a case of ideal coherence, the relationship between Present Hortense and Future Tuesday Hortense would look like, Present Hortense’s error ceases to appear *prudential* at all. Hortense knows full well that the Future Tuesday stage of herself is going to suffer horribly. Her mistake isn’t any failure to register that fact; her mistake, if she is making one, is one of cruelty and indifference. She stares mercilessly at that future suffering and just doesn’t care. Hortense’s preference for pain on Future Tuesdays may well be intrinsically irrational, in other words, but if so, it now seems that the mistake being made is moral. Once we notice this, however, putting these characters in the same body and introducing the complication of Tuesdays seems more of a distraction than anything. So let’s put the torturer back outside the body of the victim and turn to the case of the ideally coherent Caligula. If my treatment of Future Tuesday Indifference has been successful, moral intuitions are all that remain of one’s original intuitions of irrationality about the case.

From the case of Hortense, one quick general lesson:

Lesson 6: Examples appealing to a person’s relationship with herself over time require special care. We human beings tend to have a certain kind of relationship of caring with our future selves. In testing one’s intuitions about the normative reasons of ICEs, it is important not implicitly to be assuming that the character shares those ordinary cares simply in virtue of having been stipulated to be the same person over time. If we imagine someone who in full coherence and making no non-normative mistakes whatsoever shows absolutely no regard for the interests of his future self at a given time or set of times, then we may discover that the issues in play are actually more ones of morality than prudence.

4. Third case study: The ideally coherent Caligula

One might be willing to grant that a weird enough character could have most normative reason to starve herself to death for the sake of a “trim” figure, or to choose an agonizing operation on a Tuesday over a mildly painful one the next day. But one might think the situation is entirely different when it comes to ICEs who see no reason not to inflict suffering on others.

Regarding Gibbard's useful example of an ideally coherent Caligula,⁴⁴ for example, one's intuitions might seem much more unshakable than no matter how consistent and perfectly informed an idealized version of Caligula is, it is simply false that maximizing the suffering of others is what this man has most normative reason to do.

Amoral ICEs raise a large set of issues, and ultimately require a paper of their own; here there is space to give only a brief sketch of how I think these cases are best handled. In this section I'll merely state baldly, without arguing for, the additional lessons about ICEs that I think would emerge from a careful consideration of an ideally coherent Caligula.⁴⁵

Lesson 7: When testing one's intuitions about amoral ICEs for the sake of deciding between attitude-dependent and attitude-independent conceptions of value, special care is required to eliminate interference from irrelevant linguistic intuitions in the direction of what Stephen Darwall calls *morality/reasons externalism*.⁴⁶ Intuitions in the direction of morality/reasons externalism, roughly, are intuitions according to which the conclusion that something is "morally required" doesn't yet settle whether a person has "normative reason" to do it. In short, when thinking about the ideally coherent Caligula, we need to make sure that we're not merely asking what Caligula has most *moral* reason to do, but rather what he has most reason to do *full stop*. For of course refraining from torture is what Caligula has most *moral* reason to do; one might think this is close to a conceptual truth about what morality consists in, since any "morality" that gave a green light to maximizing the suffering of others for fun would be hard to recognize as a system of *morality* at all. But the question we're interested in is whether Caligula has most normative reason *full stop* to do what he has most "moral reason" to do. In other words, of course torturing people for fun is *immoral*; but the question we're interested in is whether an ideally coherent Caligula should (full stop) be moral. Talk about anything else, and you're not talking about normativity, though it might sometimes sound as though you are.⁴⁷

Lesson 8: Don't forget that an attitude-dependent theorist may in fact be able to accommodate the idea that Caligula has most normative reason not to try to maximize the suffering of others. This would be so if the Kantian project—at least on some interpretations, such as that of Korsgaard—succeeds in showing how if one values anything at all, it is entailed (albeit in a non-obvious way) from those values, whatever they may be, that one has most normative reason to act morally.⁴⁸ On this view, maximizing the suffering of others is *not* a coherent life policy, contrary to what it might seem on superficial inspection. I am pessimistic that this Kantian project can be made to succeed, but the point here is just that in debates about the attitude-dependence of value, supporters of attitude-independent conceptions cannot merely assume without argument that this way for the attitude-dependent theorist to try to deal with Caligula doesn't succeed.⁴⁹

Lesson 9: In thinking about amoral ICEs—and assuming now for the sake of argument that the Kantian project doesn't succeed—it is important to remember all the things an attitude-dependent theorist can say about them and regard herself as speaking truly.⁵⁰ For instance, we can say that we loathe the ideally coherent Caligula; that it's awful, from our point of view, that he and his normative reasons are like this; and that the rest of us who do care about morality have every normative reason to lock him up, defend ourselves against him, and to try to change him if we can.

Lesson 10: In thinking about amoral ICEs—and again assuming the Kantian project doesn't succeed—it is also important to remember all the things an attitude-dependent theorist can say about them and regard herself as speaking falsely but with good reason. "Browbeating" has a bad name,⁵¹ but sometimes using language as a causal influence, rather than to state the truth, is what we have most normative reason to do. So we might be perfectly justified in saying to the ideally coherent Caligula and to each other that Caligula has most normative reason not to

torture people even if we think this is strictly speaking false. This might be justified in roughly the same way one might be justified in telling a cancer patient that she is looking strong and well even if one regards this as strictly speaking false. What matters in such cases (to those of us who care about others) is what one hopes to do or encourage or express with one's words, not what is strictly speaking true.

Lesson 11: When it comes to amoral ICEs, the analogy of alien creatures is especially important to keep in mind. If a group of intelligent, ideally coherent aliens descended upon us and began trying to kill us for food or torture us for sport, would we feel intuitively convinced that they were making a mistake about the normative facts? Would we be more inclined to say "They shouldn't be doing this" or rather just "How can we stop them?" When it comes to real-life people, those of us deeply committed to morality should always err in the direction of assuming that an alleged amoralist is making a mistake, for in real life such a person is in all likelihood inconsistent, wrong about many of the non-normative facts, and self-deceived. Most real-life human beings have moral feelings that can be tapped into, and in talking to them and others about their normative reasons we can appeal to that. The more we start to suspect that a person hasn't even the tiniest trace of moral feeling, however, the more the alien species analogy will start to seem the right one.

Lesson 12: Never forget the holistic nature of the debate about the attitude-dependence of value. Always driving the attitude-dependent theorist's willingness to bite certain bullets, including the bullet that an ideally coherent Caligula does not have most normative reason to act morally, is the conviction that this conclusion is ultimately forced upon us by metaphysical and epistemological considerations and more generally a naturalistically comprehensible understanding of the world. In my own view, the price of cleaving to an attitude-independent conception is nothing less than total skepticism about our ability to grasp the normative truth.⁵² In my view, this result would be unacceptable and settle the debate in favor of attitude-dependent conceptions even if claims (1')-(5') were extremely implausible. But as I've tried to show in this paper, this isn't so: claims (1')-(5') turn out not to be implausible at all.

5. Conclusion

There is a sense in which nothing hinges on what we say about ICEs, and there is a sense in which everything does. Nothing hinges on these cases in the sense that none of us will ever actually encounter an ICE. But everything hinges on these cases in the sense that our whole conception of ourselves and what we're doing when we engage in practical reasoning depends on the right thing to say about them. How should we understand the relationship between ourselves, value, and the world? When we face the deepest, toughest questions about how to conduct our lives, is the kind of fidelity we're after fidelity to something independent or is it fidelity to ourselves? We encounter ICEs in our imaginations, and that's enough: we are forced by their mere possibility to ask whether there is something they would be missing that we're not. If we were like them rather than ourselves, in other words, would we be in error about something, or would we just be different?

There is a widespread habit in metaethics of citing ideally coherent eccentrics with only cursory discussion and assuming that one has thereby established or substantially bolstered the case for an attitude-independent conception of normativity. It is time to put an end to that habit. Those who think that there are attitude-independent facts about normative reasons are wrong if they think they can depend upon these odd characters to do their philosophical heavy lifting.⁵³

Notes

¹ Hume 1739-40, bk. 2, pt. 3, sec. 3.

² Rawls 1971, p. 432.

³ Gibbard 1990, p. 171.

⁴ Gibbard 1999, p. 145.

⁵ Parfit 1984, p. 124, and forthcoming, ch. 2, sec. 7.

⁶ Hume and Rawls come up with their examples in the course of *denying* this view. But many philosophers have regarded Hume and Rawls as biting implausible bullets in these passages, so the characters in question are frequently invoked to motivate the very view Hume and Rawls were denying.

⁷ Nothing important hinges on my use of reasons terminology in this paper. While it's my language of choice for talking about normativity, anyone with reservations about "reasons talk" is welcome to substitute his or her own language of choice. All that matters is that we are talking about the subject matter of *what to do, full stop*. For further discussion, see Street 2008b, especially sec. 8.

⁸ I use *evaluative attitudes* as a rough blanket term to cover an agent's values, desires, states of approval and disapproval, unreflective evaluative tendencies, and consciously or unconsciously held normative judgments.

⁹ Examples of theorists who hold attitude-independent conceptions include Dworkin 1996, Enoch 2007, FitzPatrick 2008, Huemer 2005, Nagel 1986, Parfit 1984 and forthcoming, Scanlon 1998, and Shafer-Landau 2003. While the matter is complicated and not something there is space to discuss here, there is also an important sense in which quasi-realists such as Blackburn and Gibbard accept an attitude-independent conception of normative reasons, though in their case it is crucial to note that they understand the position as a substantive normative claim as opposed to a metaethical position. See Street 2009b for further discussion of quasi-realism. The *attitude-independent/attitude-dependent* terminology is not ideal, but other terminological options in the ballpark are misleading. *Realism/antirealism* is misleading because the debate I'm interested in may be viewed as a substantive normative debate in which expressivists such as Blackburn and Gibbard endorse an *attitude-independent* view; *objectivism/subjectivism* is misleading because *attitude-dependent* views are in principle compatible with an extremely strong form of objectivity (for example, if arguments such as Korsgaard's in 1996b succeed); and *externalism/internalism* is misleading because of its distracting associations with the debate sparked by Williams 1981. While my sympathies are very much with Williams's view, my own objections to the idea of attitude-independent normative reasons are mainly epistemological, and not of the kind Williams raises in his 1981.

¹⁰ The view defended in Williams 1981 is a paradigmatic example of an attitude-dependent conception. The "formalist" or "Humean" version of constructivism that I defend in Street 2008a and discuss further in Street 2009a is another example. The view defended in Korsgaard 1996b is one of the most prominent contemporary examples of an attitude-dependent conception, but her view differs from Williams's and mine in that Korsgaard believes that strong substantive normative conclusions—specifically, a commitment to "Enlightenment morality"—are entailed from within the standpoint of any set of values whatsoever.

¹¹ For suggestive discussion of how killing oneself for the sake of a contingent end might involve an incoherence, see Korsgaard's discussion of Kant's views on suicide in her 1996a. For an attempt to show that immorality involves a kind of incoherence, see Korsgaard 1996b.

¹² I return to it briefly in section 4 when I list "Lesson 8."

¹³ There are many important complexities here which there is not space to discuss. As "Lesson 10" in section 4 suggests, for example, in real-life interactions, it can be perfectly appropriate to say things that one regards as strictly speaking false.

¹⁴ Street 2006, 2008b, 2009b, and 2009c.

¹⁵ I read Dworkin as making exactly this point in his 1996, pp. 117-118. For further discussion, see Street 2009c, section 12.

¹⁶ For a parade of philosophically interesting ICEs, see Parfit 1984, pp. 124-126, and Parfit forthcoming, chs. 2-4.

¹⁷ For those who think it is a mistake to dwell on outlandish hypothetical cases such as those of ICEs, I address methodological questions midway through section 2 and in the conclusion.

¹⁸ Gibbard 1990, p. 171.

¹⁹ *Ibid.*, p. 166.

²⁰ Here I am glossing over many important complexities in Gibbard's position. I discuss expressivism and quasi-realism in 2009b.

²¹ The language of earning or preserving our "right" to make objectivist claims about normative matters comes from Blackburn; see his 1984, p. 197.

²² Gibbard, 1990, p. 155 and p. 164.

²³ See Street 2009b for a more detailed discussion of these positions.

²⁴ Gibbard himself rightly expresses qualms about using the anorexia example, emphasizing in a note that "real anorexia nervosa does not consist in acting on coherent preferences" (1990, p. 165). As I'll try to show, though, the need for caution with the example is significantly greater than even Gibbard's own strong expression of qualms recognizes.

²⁵ Anorexia afflicts men as well as women. Because the disease afflicts women disproportionately, however, and because Gibbard's character is female, I focus on women in my discussion.

²⁶ I want to emphasize again (see note 24) that *Gibbard* does not make the mistake of thinking that real-life anorexia ever involves ideal coherence. Still, I think there is a way in which the full importance of this point gets lost in his discussion.

²⁷ For a sketch of how I think the alleged "values" of logical and instrumental consistency can be explained, see Street 2008a, especially section 7.

²⁸ Parfit forthcoming, ch. 2, sec. 7.

²⁹ *Ibid.*

³⁰ Parfit 1984, p. 125.

³¹ Parfit forthcoming, chs. 2-4.

³² For further discussion of the nature of pain and its reason-giving status, see Street 2006, sec. 9.

³³ Parfit forthcoming, ch. 2, sec. 6.

³⁴ Ibid.

³⁵ Parfit 1984, p. 124.

³⁶ Kabat-Zinn 1994, p. 141.

³⁷ I have in mind the case of Thich Quang Duc and others like him.

³⁸ I am ignoring complications about time zone, daylight savings, and so forth. While the story about causal origins would have to be even more bizarre to accommodate these complications due to the conventional nature of Tuesdays, I don't think anything of philosophical significance ultimately hinges on the matter. Thanks to Gilbert Harman and Peter Unger for discussion of this.

³⁹ One might object that "pains" not accompanied by *meta-hedonic* distress aren't really pains; accordingly, one might argue that Indy, due to his meta-hedonic indifference, isn't genuinely in pain when he undergoes the tooth extraction. My reply to this is twofold. First, it just seems wrong: Indy is in terrible bodily pain in the same way that a monk burning himself to death or a woman giving birth is in terrible bodily pain even if he or she is simultaneously experiencing meta-hedonic calm (in the monk's case) or joy (in the woman's). It is simply very implausible to claim that someone who is burning to death or giving birth is not genuinely in serious bodily pain, even if at a meta-hedonic level he or she is doing just fine. Second, my main point doesn't hinge on this anyway. If you think that a "pain" isn't a pain unless it's accompanied by *meta-hedonic* distress, then you already agree with me that Indy's case fails to provide an example of an intrinsically irrational preference. You'll be more interested in the case of Hortense, which I turn to next.

⁴⁰ For further discussion of pain, see Street 2006, sec. 9.

⁴¹ Parfit 1984, pp. 124-125.

⁴² Parfit 1984, p. 124.

⁴³ Ibid.

⁴⁴ Gibbard 1999, p. 145.

⁴⁵ I say a little bit more about these issues in Street 2009c.

⁴⁶ Darwall 1997, p. 306.

⁴⁷ I say more about these issues in Street 2008b.

⁴⁸ Korsgaard 1996b.

⁴⁹ Gibbard in his 1999 does *not* merely assume without argument that this Kantian route doesn't succeed; he explicitly considers it and is skeptical.

⁵⁰ Cf. Williams 1995, p. 39, who writes that "There are many things I can say about or to [a man whose motivational set gives him no reason to be nicer to his wife]: that he is ungrateful, inconsiderate, hard, sexist, nasty, selfish, brutal, and many other disadvantageous things." Williams is not careful enough about his choice of language in such passages, since terms such as *selfish* and *brutal* carry with them the strong implication that the person so described has normative reason to be different. On this point, see Scanlon 1998, p. 367.

⁵¹ See Scanlon 1998, p. 365, discussing Williams's arguments against externalism.

⁵² Street 2006, 2008b, 2009b, and 2009c.

⁵³ For their helpful feedback on earlier versions of this paper, I am indebted to audiences at the University of California at Berkeley, CUNY Graduate Center, Johns Hopkins University, Princeton University, and the University of South Carolina at Columbia. For their detailed comments, I am especially grateful to Richard Chappell, Elizabeth Harman, Derek Parfit, Steven Ross, and David Velleman.

References

- Blackburn, Simon. 1984. *Spreading the Word: Groundings in the Philosophy of Language*. Oxford: Clarendon Press.
- Darwall, Stephen. 1997. "Reasons, Motives, and the Demands of Morality: An Introduction." In *Moral Discourse and Practice: Some Philosophical Approaches*, ed. Darwall, Gibbard, and Railton. New York: Oxford University Press.
- Dworkin, Ronald. 1996. "Objectivity and Truth: You'd Better Believe It." *Philosophy & Public Affairs* 25: 87-139.
- Enoch, David. 2007. "An Outline of an Argument for Robust Metanormative Realism." *Oxford Studies in Metaethics*, vol. 2, ed. Russ Shafer-Landau. Oxford: Clarendon Press.
- FitzPatrick, William. 2008. "Robust Ethical Realism, Non-Naturalism, and Normativity." *Oxford Studies in Metaethics*, vol. 3, ed. Russ Shafer-Landau. Oxford: Clarendon Press.
- Gibbard, Allan. 1990. *Wise Choices, Apt Feelings*. Cambridge, MA: Harvard University Press.
- Gibbard, Allan. 1999. "Morality as Consistency in Living: Korsgaard's Kantian Lectures." *Ethics* 110: 140-164.
- Gibbard, Allan. 2003. *Thinking How to Live*. Cambridge, MA: Harvard University Press.
- Huemer, Michael. 2005. *Ethical Intuitionism*. Palgrave Macmillan.
- Hume, David. 1739-40. *A Treatise of Human Nature*, second edition, eds. L. A. Selby-Bigge and P. H. Nidditch. Oxford: Oxford University Press, 1978.
- Kabat-Zinn, Jon. 1995. *Wherever You Go, There You Are*. New York: Hyperion, 1994.
- Korsgaard, Christine. 1996a. *Creating the Kingdom of Ends*. Cambridge: Cambridge University Press.
- Korsgaard, Christine. 1996b. *The Sources of Normativity*. Cambridge: Cambridge University Press.
- Nagel, Thomas. 1986. *The View From Nowhere*. Oxford: Oxford University Press.
- Parfit, Derek. 1984. *Reasons and Persons*. Oxford: Oxford University Press.
- Parfit, Derek. Forthcoming. *On What Matters*. Oxford: Oxford University Press.
- Rawls, John. 1971. *A Theory of Justice*. Cambridge, MA: Harvard University Press.
- Scanlon, T. M. 1998. *What We Owe to Each Other*. Cambridge, MA: Harvard University Press.
- Shafer-Landau, Russ. 2003. *Moral Realism: A Defence*. Oxford: Clarendon Press.
- Street, Sharon. 2006. "A Darwinian Dilemma for Realist Theories of Value." *Philosophical Studies* 127: 109-166.

- Street, Sharon. 2008a. "Constructivism About Reasons." *Oxford Studies in Metaethics*, vol. 3, ed. Russ Shafer-Landau. Oxford: Clarendon Press.
- Street, Sharon. 2008b. "Reply to Copp: Naturalism, Normativity, and the Varieties of Realism Worth Worrying About." *Philosophical Issues* (a supplement to *Noûs*), vol. 18 on "Interdisciplinary Core Philosophy," ed. Walter Sinnott-Armstrong.
- Street, Sharon. 2009a. "What is Constructivism in Ethics and Metaethics?" Forthcoming in *Philosophy Compass*, vol. 4, 2009.
- Street, Sharon. 2009b. "Mind-Independence Without the Mystery: Why Quasi-Realists Can't Have It Both Ways." Forthcoming.
- Street, Sharon. 2009c. "Objectivity and Truth: You'd Better Rethink It." Forthcoming.
- Williams, Bernard. 1981. "Internal and External Reasons." In *Moral Luck*. Cambridge: Cambridge University Press.
- Williams, Bernard. 1995. "Internal Reasons and the Obscurity of Blame." In *Making Sense of Humanity*. Cambridge: Cambridge University Press.